JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 10 (2022), Issue 3 ISSN 2291-9279 Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Designing Virtual Reality–Based Conversational Agents to Train Clinicians in Verbal De-escalation Skills: Exploratory Usability Study (e38669)
Nathan Moore, Naseem Ahmadpour, Martin Brown, Philip Poronnik, Jennifer Davids. 5

Effect of the “Art Coloring” Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation (e37026)
JuZhe Xi, YuHan Gao, Na Lyu, Zhuang She, XinYue Wang, Xin-An Zhang, XiaoYu Yu, WeiDong Ji, MengSheng Wei, WeiHui Dai, Xuesheng Qian. 18

The Long-term Effects of Immersive Virtual Reality Reminiscence in People With Dementia: Longitudinal Observational Study (e36720)
Ling-Chun Huang, Yuan-Han Yang. 49

Social Media Users’ Perceptions of a Wearable Mixed Reality Headset During the COVID-19 Pandemic: Aspect-Based Sentiment Analysis (e36850)
Heejin Jeong, Allison Bayro, Sai Umesh, Kaushal Mamgain, Moontae Lee. 68

Breathing as an Input Modality in a Gameful Breathing Training App (Breeze 2): Development and Evaluation Study (e39186)
Yanick Lukic, Gisbert Teepe, Elgar Fleisch, Tobias Kowatsch. 85

Development of a Therapeutic Video Game With the MDA Framework to Decrease Anxiety in Preschool-Aged Children With Acute Lymphoblastic Leukemia: Mixed Methods Approach (e37079)
Dai-Jie Yang, Meng-Yao Lu, Chi-Wen Chen, Pei-Ching Liu, I-Ching Hou. 106

Cognitive Behavioral Therapy Plus a Serious Game as a Complementary Tool for a Patient With Parkinson Disease and Impulse Control Disorder: Case Report (e33858)
Teresa Mena-Moreno, Lucero Munguía, Rosario Granero, Ignacio Lucas, Almudena Sánchez-Gómez, Ana Cámara, Yaroslau Compta, Francesc Valldeoriola, Fernando Fernandez-Aranda, Anne Sauvaget, José Menchón, Susana Jiménez-Murcia. 120

Prediction of Specific Anxiety Symptoms and Virtual Reality Sickness Using In Situ Autonomic Physiological Signals During Virtual Reality Treatment in Patients With Social Anxiety Disorder: Mixed Methods Study (e38284)
Joo Chun, Hyun-Jin Kim, Ji-Won Hur, Dooyoung Jung, Heon-Jeong Lee, Seung Pack, Sungkil Lee, Gerard Kim, Chung-Yean Cho, Seung-Moo Lee, Hyeri Lee, Seungmoon Choi, Taesu Cheong, Chul-Hyun Cho. 150

The Effects of a Computer Game (Healthy Rat King) on Preschool Children’s Nutritional Knowledge and Junk Food Intake Behavior: Nonrandomized Controlled Trial (e33137)
Ing-Chau Chang, Cheng-Ying Yang, Chin-En Yen. 165

XSL•FO
RenderX

Preliminary User Evaluation of a New Dental Technology Virtual Simulation System: Development and Validation Study (e36079)
Mengwei Pang, Xiaohan Zhao, Daiyu Lu, Yihan Dong, Lin Jiang, Jie Li, Ping Ji. 174

Migration of an Escape Room–Style Educational Game to an Online Environment: Design Thinking Methodology (e32095)
Maja Videnovik, Tone Vold, Georgina Dimova, Linda Kiønig, Vladimir Trajkovik. 193

Design of a Game-Based Training Environment to Enhance Mental Health Care Professionals' Skills in Using e–Mental Health: Multiple Methods User Requirements Analysis (e34700)
Joyce Bierbooms, Milou Feijt, Wijnand IJsselsteijn, Inge Bongers. 206

A Biofeedback-Based Mobile App With Serious Games for Young Adults With Anxiety in the United Arab Emirates: Development and Usability Study (e36936)
Mariam Almeqbaali, Sofia Ouhbi, Mohamed Serhani, Leena Amiri, Reem Jan, Nazar Zaki, Ayman Sharaf, Abdulla Al Helali, Eisa Almheiri. . .
2 2 0

Digital Biomarkers for Well-being Through Exergame Interactions: Exploratory Study (e34768)
Despoina Petsani, Evdokimos Konstantinidis, Aikaterini-Marina Katsouli, Vasiliki Zilidou, Sofia Dias, Leontios Hadjileontiadis, Panagiotis Bamidis. 233

The Effects of Acute Virtual Reality Exergaming on Mood and Executive Function: Exploratory Crossover Trial (e38200)
Genta Ochi, Ryuta Kuwamizu, Tomomi Fujimoto, Koyuki Ikarashi, Koya Yamashiro, Daisuke Sato. 248

The Impact of Cognitive Function on Virtual Reality Intervention for Upper Extremity Rehabilitation of Patients With Subacute Stroke: Prospective Randomized Controlled Trial With 6-Month Follow-up (e33755)
Yan Leng, Wai Lo, Yu Mao, Ruihao Bian, Jiang Zhao, Zhiqin Xu, Le Li, Dong Huang. 261

Interactive Somatosensory Games in Rehabilitation Training for Older Adults With Mild Cognitive Impairment: Usability Study (e38465)
Chien-Hsiang Chang, Chung-Hsing Yeh, Chien-Cheng Chang, Yang-Cheng Lin. 273

Virtual Reality as a Therapy Tool for Walking Activities in Pediatric Neurorehabilitation: Usability and User Experience Evaluation (e38509)
Corinne Ammann-Reiffer, Andrina Kläy, Urs Keller. 291

The Design, Development, and Testing of a Virtual Reality Device for Upper Limb Training in People With Multiple Sclerosis: Single-Center Feasibility Study (e36288)
Alon Kalron, Lior Frid, Iliya Fonkatz, Shay Menascu, Mark Dolev, David Magalashvili, Anat Achiron. 301

Virtual Reality–Based Rehabilitation as a Feasible and Engaging Tool for the Management of Chronic Poststroke Upper-Extremity Function Recovery: Randomized Controlled Trial (e37506)
Alejandro Hernandez, Liudmila Bubyr, Philippe Archambault, Johanne Higgins, Mindy Levin, Dahlia Kairy. 309

Speech Processing as a Far-Transfer Gauge of Serious Games for Cognitive Training in Aging: Randomized Controlled Trial of Web-Based Effectivate Training (e32297)
Gal Nitsan, Shai Baharav, Dalith Tal-Shir, Vered Shakuf, Boaz Ben-David. 322

Virtual ER, a Serious Game for Interprofessional Education to Enhance Teamwork in Medical and Nursing Undergraduates: Development and Evaluation Study (e35269)
Janet Wong, Joanna Ko, Sujin Nam, Tyrone Kwok, Sheila Lam, Jessica Cheuk, Maggie Chan, Veronica Lam, Gordon Wong, Zoe Ng, Abraham Wai. 337

XSL•FO
RenderX

An Augmented Reality–Based Guide for Mechanical Ventilator Setup: Prospective Randomized Pilot Trial (e38433)
Sejin Heo, Suhyeon Moon, Minha Kim, Minsu Park, Won Cha, Meong Son. . . . 349

Points and the Delivery of Gameful Experiences in a Gamified Environment: Framework Development and Case Analysis (e35907)
Sungjin Park, Sangkyun Kim. . . . 359

A New Approach for Reducing Virtual Reality Sickness in Real Time: Design and Validation Study (e36397)
JuHye Won, Yoon Kim. . . . 371

Effectiveness of Using Augmented Reality for Training in the Medical Professions: Meta-analysis (e32715)
Yahia Baashar, Gamal Alkawsi, Wan Ahmad, Hitham Alhussian, Ayed Alwadain, Luiz Capretz, Areej Babiker, Adnan Alghail. . . . 383

Reviews

Effectiveness of Serious Games for Improving Executive Functions Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis (e36123)
Alaa Abd-alrazaq, Dari Alhuwail, Arfan Ahmed, Mowafa Househ. . . . 31

The Effectiveness of Serious Games on Cognitive Processing Speed Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis (e36754)
Alaa Abd-alrazaq, Arfan Ahmed, Haitham Alali, Ahmad Aldardour, Mowafa Househ. . . . 135

Effect of Virtual Reality on Pediatric Pain and Fear During Procedures Involving Needles: Systematic Review and Meta-analysis (e35008)
Marta Lluesma-Vidal, Raquel Carcelén González, Laura García-Garcés, María Sánchez-López, Loreto Peyro, Cayetana Ruiz-Zaldibar. . . . 396

The Effectiveness of Serious Games in Improving Memory Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis (e35202)
Alaa Abd-alrazaq, Dari Alhuwail, Eiman Al-Jafar, Arfan Ahmed, Farag Shuweihdi, Shuja Reagu, Mowafa Househ. . . . 408

Digital Interventions for Emotion Regulation in Children and Early Adolescents: Systematic Review and Meta-analysis (e31456)
Sally Reynard, Joao Dias, Marija Mitic, Beate Schrank, Kate Woodcock. . . . 433

Serious Games Supporting the Prevention and Treatment of Alcohol and Drug Consumption in Youth: Scoping Review (e39086)
Juan Martínez-Miranda, Ismael Espinosa-Curiel. . . . 457

Augmented Reality in Vascular and Endovascular Surgery: Scoping Review (e34501)
Joshua Eves, Abhilash Sudarsanam, Joseph Shalhoub, Dimitri Amiras. . . . 480

Tutorial

Developing Serious Video Games to Treat Attention Deficit Hyperactivity Disorder: Tutorial Guide (e33884)
Aarón Sújar, Marina Martín-Moratinos, María Rodrigo-Yanguas, Marcos Bella-Fernández, Carlos González-Tardón, David Delgado-Gómez, Hilario Blasco-Fontecilla. . . . 58

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Effect of the “Art Coloring” Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation (e41253)
JuZhe Xi, YuHan Gao, Na Lyu, Zhuang She, XinYue Wang, Xin-An Zhang, XiaoYu Yu, WeiDong Ji, MengSheng Wei, WeiHui Dai, XueSheng Qian 381

XSL•FO
RenderX

Original Paper

Designing Virtual Reality–Based Conversational Agents to Train Clinicians in Verbal De-escalation Skills: Exploratory Usability Study

Nathan Moore[1], BN, MEd; Naseem Ahmadpour[2], PhD; Martin Brown[3], MH; Philip Poronnik[3], PhD; Jennifer Davids[4], PhD

[1]Digital Health Solutions, Western Sydney Local Health District, North Parramatta, Australia

[2]Design Lab, Sydney School of Architecture, Design and Planning, The University of Sydney, Sydney, Australia

[3]Faculty of Medicine and Health Media Lab, Education Innovation, School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[4]Research and Education Network, Western Sydney Local Health District, Westmead, Australia

Corresponding Author:
Nathan Moore, BN, MEd
Digital Health Solutions
Western Sydney Local Health District
Cumberland Hospital East Campus
Building 106
North Parramatta, 2151
Australia
Phone: 61 0427850889
Email: nathan.moore@health.nsw.gov.au

Abstract

Background: Violence and aggression are significant workplace challenges faced by clinicians worldwide. Traditional methods of training consist of “on-the-job learning” and role-play simulations. Although both approaches can result in improved skill levels, they are not without limitation. Interactive simulations using virtual reality (VR) can complement traditional training processes as a cost-effective, engaging, easily accessible, and flexible training tool.

Objective: In this exploratory study, we aimed to determine the feasibility of and barriers to verbal engagement with a virtual agent in the context of the Code Black VR application. Code Black VR is a new interactive VR-based verbal de-escalation trainer that we developed based on the Clinical Training Through VR Design Framework.

Methods: In total, 28 participants with varying clinical expertise from 4 local hospitals enrolled in the Western Sydney Local Health District Clinical Initiative Nurse program and Transition to Emergency Nursing Programs and participated in 1 of 5 workshops. They completed multiple playthroughs of the Code Black VR verbal de-escalation trainer application and verbally interacted with a virtual agent. We documented observations and poststudy reflection notes. After the playthroughs, the users completed the System Usability Scale and provided written comments on their experience. A thematic analysis was conducted on the results. Data were also obtained through the application itself, which also recorded the total interactions and successfully completed interactions.

Results: The Code Black VR verbal de-escalation training application was well received. The findings reinforced the factors in the existing design framework and identified 3 new factors—motion sickness, perceived value, and privacy—to be considered for future application development.

Conclusions: Verbal interaction with a virtual agent is feasible for training staff in verbal de-escalation skills. It is an effective medium to supplement clinician training in verbal de-escalation skills. We provide broader design considerations to guide further developments in this area.

(JMIR Serious Games 2022;10(3):e38669) doi:10.2196/38669

KEYWORDS

virtual reality; code black; verbal de-escalation; violence and aggression; education; clinical training; conversational agent

XSL•FO
RenderX

Introduction

Background

Violence and aggression in health care settings is an international problem, with one-fifth of health care professionals experiencing violence perpetrated by patients or family members every year [1]. The implementation of effective training in high-quality verbal de-escalation skills has been shown to be vital for the well-being of staff and patients. When used early, these skills can prevent escalation of the situation to violence [2].

"A Code Black is any incident where hospital staff are threatened with or experience verbal, physical or psychological abuse or attack during the course of their employment and a staff response is required" [3]. The code black response involves a variety of interventions that may be required preincident occurrence, during the incident, or postincident occurrence [3]. There is evidence that the delivery of high-quality violence and aggression minimization training has a positive impact on staff member's perceived confidence in managing these situations [4]. In addition, this style of training corelates with a reduction in the number of incidents of violence and aggression occurring in health care [5].

Current practice for Violence Prevention and Management training within our Local Health District (LHD) is conducted using a traditional face-to-face didactic simulation and skills station format. This training allows clinicians to collaborate and practice these skills in a group setting with expert support from course instructors. Simulation has been clearly established as an effective way to train clinicians in the application of skills and to reduce anxiety before the application of these skills [6]. However, a significant challenge posed by this type of training is that it requires significant resources and clinicians to be present in a fixed location and time [7]. Given the resource limitations, the complexity of staffing requirements, and the high number of staff members requiring training, the current demand for this training far exceeds the capacity for the program's delivery. In addition, there is a need for refresher training for staff, and this is not factored into the current program's face-to-face delivery schedule. e-Learning training modules and videos are available; however, there are questions about the efficacy of e-learning to impact health care professional's behaviors, skills, or knowledge [8].

Owing to these challenges and the increasing occurrence of code black, we developed a pilot supplemental virtual reality (VR)–based application to support clinician education in code black management. The initial phase of this project scoped the breadth of the problem and involved staff interviews and incident review. Our initial findings highlighted that staff required training on how to recognize behaviors of concern in an individual and practice in verbal de-escalation skills to minimize or prevent aggression or violence toward a staff member [2]. Therefore, we reasoned that there is a need for readily accessible and frequently repeatable experiential training in managing code black events [3].

To address the barriers of access, cost, and availability of training, VR is being adopted as a supplemental modality to support clinician education in several scenarios [9]. For example, VR has been explored to train students in conflict management as part of the research being conducted by the University of Newcastle [10]. While still in trial, the approach is being well received by students and is showing potential for providing education in this challenging area. Previously, we used VR as a tool within the health district to train clinicians in the management of advanced life support (ALS) [11]. Building upon our experience in this area and our findings in previous VR user needs analysis [12], we developed a VR-based application to support code black training.

VR Technology for Verbal De-escalation Training

The novel capabilities of VR provide significant potential for educators to use this approach to supplement and, in some cases, replace traditional learning modalities [13,14]. The interaction with 3D representations of people, items, and environments in real time can allow users to practice the application of skills. Deploying a VR app on portable and standalone VR head-mounted displays (HMDs) enables users to engage with training in a time and setting of their choosing, which may not have been possible otherwise [9,15]. In addition, the level of *blocking out* of the physical world made possible by the VR HMDs can further increase the immersion of the user and increase presence; that is, the feeling of being "present" in the scenario [16]. This results in the suspension of disbelief and the generation of authentic reactions in users due to increased engagement with the experience. The ability to record and assess user interactions within a VR app allows for completely objective, structured assessment and feedback on the platform, which can be difficult with human assessors [17]. Voice recognition within the application can add an element of increased engagement with the technology but also brings with it challenges and limitations with regard to misinterpretation or misunderstanding errors, which must be overcome [18].

Conversational Agents for Training

Increasing processor power and technological advancements have resulted in the emergence of virtual conversational agents as a potential training modality. BodySwaps is a company developing VR-based "soft skills" training where users can observe interactions between virtual agents and make structured observations and interact with the virtual agent to provide comment and feedback [19]. The responses are recorded and provided as feedback to the user for development. The content provided to the users focuses on human resources and management-style scenarios. Other studies using virtual conversational agents have been conducted by the University of Newcastle in an application built to train student nurses called "Angry Stan." The application uses biofeedback to train clinicians to remain calm in confronting situations. The user responses are chosen from a multiple-choice list of scripted responses, and the agent responds to both the choices made by the user and their chosen responses [20].

XSL•FO
RenderX

Objectives

It is clearly identified that a structured and coordinated response by well-trained staff using a shared organizational approach is vital to minimize the risk to both staff and patients [21]. Given the identified barriers to the accessibility of existing training approaches, we aimed to develop a VR-based supplemental code black training application. We have previously developed VR-based applications to support clinical education in areas such as ALS management, leadership, clinical handover, and dignity in the workplace [11]. However, verbal interaction with a virtual agent that can respond to user input is a far more complex challenge, and there is little research that can guide the design of such interactions.

The aim of this study was 2-fold: (1) to determine the validity of using the existing "Clinical Training Through VR Design Framework" to assess the feasibility of VR-based education modules [12] and (2) to identify specific design requirements for a VR-based agent to train clinicians in verbal de-escalation skills in the context of code black management.

The Clinical Training Through VR Design Framework involves 8 factors that define the clinician training needs that must be supported within the VR environment and through the interactions [12]. These factors are realistic tasks, visibility, completion, accessibility, agency, diverse input modalities, mental models, and advanced roles [12]. In this original iteration of the framework, a subgrouping of some commonly associated factors was provided.

Methods

Overview

This was a mixed methods study. Opportunistic recruitment occurred during the Western Sydney LHD (WSLHD) Transition to Emergency Nursing and Clinical Initiative Nurse programs. Five workshops were conducted during these programs, each lasting approximately 20 minutes at the Westmead Hospital Simulated Learning Environment for Clinical Training simulation center. Data were collected through qualitative observation notes by researchers and a postuse survey was completed by participants, which were put into the context of analytic data obtained from the Code Black VR system.

At the start of the study, participants were given an orientation that consisted of a brief overview of the study and instructions regarding the use of the Oculus Quest 2 headset. Following the orientation, the participants undertook multiple unguided playthroughs in VR with instructors (NM, MB, and JD) observing the group and answering any questions. Each individual playthrough lasted approximately 10 minutes. Upon completion, the participants completed a questionnaire based on the System Usability Scale (SUS) [22] and had the opportunity to add open-ended comments at the end. The SUS was used for this study because it is a widely adopted tool for the evaluation of user interfaces with a focus on usability [23]. The SUS consists of 10 distinct questions designed to evaluate the user interface. Each question was rated on a 5-point scale ranging from strongly agree to strongly disagree. The final calculation provides a score of 0 to 100 with a SUS score of more than 68 being deemed "above average" [22]:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions of this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with the system.

The researchers present during the session also documented their reflections on the overall session after the workshops. The data obtained from the SUS were consolidated, and a final score was calculated to determine the overall usability of the system. A thematic analysis was conducted on participants' open-ended comments to identify common themes and better understand the user experience of the tool. Researcher observations were summarized separately, offering insights into the identified themes. In addition, the back-end application records the success and failure of speech-to-text (STT) interpretations. These analytics were collated to provide objective data on in-app use.

Code Black VR

Design Rationale

On the basis of prior experience with developing simulation-based educational experiences in both traditional and virtual settings, we identified that any solution to support this training needed to be flexible and must address the needs of both clinicians and educators alike. In our view, to be truly flexible, the VR solution had to run on portable and untethered HMDs. This allows "free range" deployment into any environment as there is no need for any additional external computers.

To address the requirement of verbal de-escalation skill training, a realistic dialogue between the user and the virtual agent is vital. In collaboration with Frameless Interactive, we used microphone access within the applications for the Oculus Quest 2 HMD. Microphone access allows voice-to-speech transcription to occur within the VR application. Keyword analysis and categorization of the resulting text file enables the app linguistic program to produce text-based responses that are then converted by text-to-speech software using the Google text-to-speech application programming interface into credible verbal interaction with a virtual agent within the HMD. Using the 8 preidentified affordances is vital for VR design for clinician education, and this guided the initial prototype build of the Code Black VR verbal de-escalation trainer that was tested in this study.

The application was created to work on the Oculus Quest 2 HMD [24]. The portable nature, processing power, and microphone capability of the Quest 2 makes it an ideal choice

XSL•FO
RenderX

for this application. The Code Black VR verbal de-escalation training application was built on the Unreal game engine [25]. Unreal was used because it has high-level textures and editing capabilities and provides a finished product of a high visual standard. In addition, we were able to leverage developments and learnings from the ALS-SimVR [11] app to improve the cost-effectiveness of the development process.

Verbal De-escalation Trainer Walkthrough

The interactive simulation positions the user in an emergency department (ED) waiting room in front of a visibly distressed digital conversational agent named Louis. The man identifies himself as the son of a patient who is awaiting review by a physician. The man is distressed by how long they have been waiting. The user who was an ED nurse, was instructed during the orientation to press the "Y" or "B" buttons on the Quest 2 controller using a push-to-talk approach to verbally respond to the patient. Following the user response, the agent responds in either a positive or negative fashion, depending on the input. Responses categorized as compassionate resulted in a decrease in the agent's level of aggression; responses categorized as confrontational resulted in an increase in the agent's level of aggression. The responses also drive an increase or decrease in the background "frustration level," which controls the agent's animations and subsequent responses. In the prototype build, the user can also see information about the progress of the scenario, such as the current "frustration level," recorded STT results, and agent responses (Figure 1).

Following a series of verbal interactions or a set period, the user will have either increased or decreased the agent's frustration level, and the scenario will conclude depending on what the user said and how they responded. Following the end of the scenario, the user is presented with their performance data (Figure 2). These data are also recorded on the SimDash Learning Management System created by Frameless Interactive to support all of our VR-based applications. This learning management system allows for more in-depth feedback. The user can later log in to review their performance over time.

Figure 1. Virtual conversational agent with status, frustration, response, and speech-to-text result.



Figure 2. User feedback after the scenario.

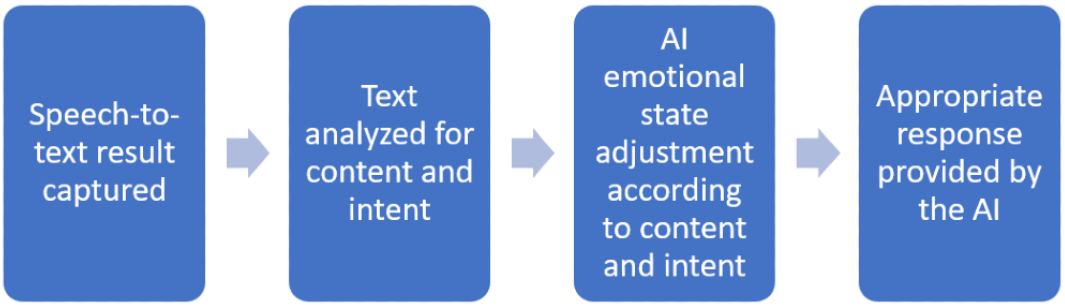


XSL•FO
RenderX

Overview of Code Black VR: Verbal De-escalation Training Application Architecture

The application places the user in a medical setting with proprietary emotional, conversational artificial intelligence (AI) called DriftAI, which drives an agent's animation and voice. At the start of the scenario, the "frustration level" of the AI, as represented by the digital agent, approaches an angry state, and the user must attempt to de-escalate the agent. The emotional state of the agent shifts depending on what the user says. The AI captures the user's speech input and analyzes the content of the conversation to determine the user's intent. It then adjusts the mood of the agent and selects an appropriate response based on the agent's emotional state for that intent (Figure 3).

Figure 3. Artificial intelligence (AI) conversation flow.



The AI response is generated in several ways, from an audio source with associated facial motion capture, AI-generated voice with lip sync, or only text. This choice depends on whether the generated response is already "known" by the system. The dialogue tree is structured in a "sandbox" format. This provides the advantage of allowing any intent to be matched at any time, reducing the need for a large number of predetermined intents to be in place. The scenario runs on a strict input and response system. The story will only advance each time the user inputs something and the agent responds to it. This is repeated until an end point is reached, such as the agent is too angry or the user has successfully de-escalated the agent.

Ethics Approval

This study was approved by the WSLHD Human Research Ethics Committee (2019/ETH00598). All the participants provided written informed consent.

Results

Overview

In total, 28 (19 female and 9 male) participants involved in the WSLHD Clinical Initiative Nurse program from 4 local hospitals participated in this study. Some participants were clinicians and educators (n=6) involved in the delivery of the program while others were junior to intermediate clinicians who were course participants (n=22). All participants had varied levels of experience working in LHD EDs, and the prerequisite of participation in the Clinical Initiative Nurse or Transition to Emergency Nursing Programs is some level of ED experience and exposure to aggression.

SUS Scoring

All collected SUS data were analyzed according to the SUS scoring procedure of adding all odd-numbered questions and subtracting 5 to obtain X, adding all even and subtracting the total from 25 to obtain Y, then adding X by Y and multiplying by 2.5 [22]. Where there was a missing response, the average of the other responses was used in the calculations. Clinicians and education groups were reported separately to aid in identifying any differences in perspective regarding perceived usability. The scores are listed in Table 1.

XSL•FO
RenderX

Table 1. System Usability Scale (SUS) scoring of the participants (N=28).

Participant	Q[a]1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS (×2.5)[b]
Clinician 1	2	4	2	4	1	0	2	4	2	4	62.5
Clinician 2	2	1	2	2	1	2	2	2	1	2	42.5
Clinician 3	3	4	3	4	3	4	3	1	3	4	42.5
Clinician 4	2	4	4	1	4	2	4	4	4	4	82.5
Clinician 5	3	4	4	1	3	1	3	1	3	3	65
Clinician 6	2	3	2	1	1	0	1	2	1	3	40
Clinician 7	2	2	3	1	2	3	3	1	3	3	57.5
Clinician 8	3	4	4	1	2	2	4	1	2	3	65
Clinician 9	3	4	0	4	3	3	4	4	3	4	80
Clinician 10	3	3	2	4	4	2	3	2	1	2	65
Clinician 11	2	2	1	0	3	DNC[c]	3	1	1	1	67[d]
Clinician 12	3	1	3	1	3	3	3	2	1	1	52.5
Clinician 13	1	4	DNC	4	2	3	3	3	2	4	72[d]
Clinician 14	4	4	4	0	3	3	4	0	4	4	75
Clinician 15	4	1	3	0	2	2	4	3	2	2	57.5
Clinician 16	3	3	4	1	2	3	4	3	2	3	65
Clinician 17	3	2	3	3	2	1	3	4	4	3	70
Clinician 18	3	4	3	3	2	2	3	3	3	3	72.5
Clinician 19	3	3	3	2	4	1	4	3	4	3	75
Clinician 20	4	2	3	0	4	2	4	1	4	2	65
Clinician 21	3	3	2	0	3	2	2	1	2	1	47.5
Clinician 22	2	1	2	2	2	2	3	2	1	1	45
Education 1	2	3	2	0	1	2	2	2	1	1	40
Education 2	1	3	2	2	1	1	2	3	1	3	47.5
Education 3	1	3	3	2	2	3	3	3	2	4	65
Education 4	1	3	1	2	2	0	1	3	DNC	3	44[d]
Education 5	3	3	3	0	2	3	3	4	3	2	65
Education 6	3	2	2	1	2	1	3	2	1	0	42.5

[a]Q: question.

[b]Average usability: 59.5.

[c]DNC: did not complete.

[d]Average of results used to calculate total.

Thematic Analysis

Three researchers (NM, MB, and JD; all experienced in educational VR development) performed a thematic analysis [26]. Owing to the clinical skill overlap between the clinical and educational participants, we decided to conduct the thematic analysis on the combined data. After coding, the comments were grouped based on the factors defined in the design framework reported in our previous study [12]. The factors identified in the previous study helped inform the development of this application and as such were relevant to the analysis of the newly gathered data [12]. The data were then reviewed using an inductive (bottom-up) approach following the process of reflexive thematic analysis (Braun and Clarke [26]) to identify any potential codes and themes that may fall outside the initial design framework.

This process helped determine whether design considerations made more broadly for VR clinical training applications correlate with participant experiences in Code Black VR given that this app uses verbal interactions with a conversational agent. Participants provided insights and comments that specifically corresponded to the 8 factors defined in our framework. However, we also identified 3 other factors that were salient in relation to Code Black VR specifically, namely motion sickness, perceived value, and privacy. For clarity, we have removed the

XSL•FO
RenderX

subcategory of affordances. All factors and exemplar descriptions are summarized in Textbox 1.

Textbox 1. The revised Clinical Training Through Virtual Reality Design Framework with 11 factors guiding the experience design and exemplar statements.

Factors and exemplar statements

- Advanced roles
  - The ability to manage tasks at an acceptable standard
- Accessibility
  - Clarity as to how commands are given and accessed
- Agency
  - The environment providing opportunity to control workflows autonomously and make choices that align with prior experiences, such as multitasking
- Completion
  - Clear commencement and completion prompt to task
- Diverse input modalities
  - The environment replicates natural input modalities such as issuing commands verbally
- Mental models
  - The environment design and prompts align with how the clinical environment operates (eg, 2 minutes of cardiopulmonary resuscitation completion)
- Motion sickness
  - That all efforts are made to reduce the experience of motion sickness for the user so they can engage with the experience
- Perceived value
  - The application provides an experience perceived as valuable by the user
- Privacy
  - The application and deployment experience should maintain the user's privacy
- Realistic tasks
  - Common clinical tasks should be available for completion in a realistic manner
- Visibility
  - Clear visible assets aligned with environmental orientation

Factors

Advanced Roles

Many participants, particularly from the educator group, identified limitations in the way the application handled some of the "higher order" or "advanced skills" they would bring to a de-escalation situation. These included considerations of physical proximity, long-form verbal responses (which were often not interpreted accurately by the application), and the ability to bring the agent's father into the discussion.

Accessibility

Participants highlighted the need for a more comprehensive orientation as to "what to do" and that it was "a bit overwhelming for a first timer." One participant provided feedback, "I would need more education on how to work it." Participants expressed preferences for the verbal interaction approach, but the "push-to-talk" method of interacting seemed to be counterintuitive or was forgotten by some participants during scenario playthroughs.

Agency

Participants wanted the ability to control and perform the required actions based on their personal experience. In the context of verbal de-escalation in health care, this poses a technological challenge. The often-lengthy verbal responses provided by clinicians were challenging for the AI to understand. This resulted in situations where participants stated, "the AI didn't respond appropriately to a lot of my responses." This often resulted in both a failure to detect what was captured accurately and interpret it correctly.

XSL•FO
RenderX

Completion

Lessons learned from our previous work in VR, which at times was unclear when the given experience was completed, informed the researchers to ensure a clear distinction for the participants as to when the verbal de-escalation scenario had finished. At the completion of the scenario, the participants were presented with feedback on their performance. They were then given the opportunity to restart the scenario and exit the application.

Diverse Input Modalities

Given the nature of verbal de-escalation skills requiring verbal interactions, participants found verbal interactions with the agent to be an important aspect of the application. Challenges were noted at times with inaccurate interpretations of what the participants had said by the application. Other input modalities, such as reducing proximity to the patient as a de-escalation tool, were also highlighted by the participants.

Mental Models

Several participants, particularly from the educator group, commented that the approach they were required to take for success within the application did not align with the approach they would take in the clinical environment. As noted in the *Advanced Roles* section, clinicians will use a variety of strategies to de-escalate a situation. The requests for response to "proximity" and being able to "utilise the dad to de-escalate the situation," all spoke to the need for greater representation of participants' mental models, which can be addressed in future iterations of the application.

Motion Sickness

Of the 28 participants, 3 (10%) either had their experience cut short or impacted due to motion sickness. One responder stated, "made me feel dizzy and I could not continue," and another stated, "I felt uneasy standing."

Perceived Value

As adult learners, for clinicians to engage with any educational approach, they must perceive the experience to have value and provide motivation to learn [27]. The perceived value of the verbal de-escalator experienced was mixed within the group, from positive sentiments, "I think it is a very clever concept and could definitely help to improve aggression in a waiting room and to learn how to de-escalate the situation before it gets worse" and "I think and can see a system like this could help train staff with de-escalation," to those who questioned its value over simulation or its value for money, "it seems like a very expensive educational tool."

Privacy

Some participants expressed concerns around "privacy." For some participants, being aware of the presence of colleagues around them participating in the same experience was confronting and challenging at times. One participant reported, "I freaked out not knowing who was near me and where," and another stated, "It felt very disconcerting having other people in the room I could hear but not see."

Realistic Tasks

The participants in our study identified a clear need to be able to complete verbal de-escalation in the virtual environment in a manner that replicated the clinical environment. These include moving further from or closer to the patient and using broad and complex language constructs. In addition, the use of emotional and facial cues from the agent were requested as they are key skills used in de-escalation.

Visibility

Participants highlighted, on multiple occasions, the need for increased visibility and realism of the agent they were de-escalating and the surrounding environments. For example, the facial expressions of the agent did not adequately reflect participants' expectations of a verbally aggressive person. One clinician stated, "his expression and body language didn't change." This is largely a technological and budgetary issue.

Observations

Observations made by the research team highlighted several barriers that inhibited the participants' ability to reach the desired objectives within the application.

Despite initial concerns about the learning curve required to access the application, based on previous VR deployments, we noted that the participants seemed to orientate to the interaction within a few minutes and engage quite quickly and naturally with the virtual agent. We suggest that this was due to the simplicity of the design; that is, there were only 2 buttons on the controller required to engage with the application (front trigger for selection and the top Y or B button for push to talk) and then interact verbally, which is an inherent skill of all participants.

We noted comments from the participants requesting training in the skills required for de-escalation before using the application. We plan to implement a training mode within the application itself that would be completed before experiencing the simulation scenario. We initially assumed that participants would feel adequately prepared to respond to the situation in this setting. However, our observations have further reinforced the need to provide the opportunity for both tutorials on how to use the application and for training for the specific skills to be used before the virtual simulation within the application for use, if required. This lack of appropriate orientation affected the accessibility of the application.

As this is a beta pilot version, the application presents the participant with minimal preparation for the scenario. However, this setting represents a common and realistic situation faced by clinicians working in the ED. Nevertheless, it was clear that this was an issue for some participants who were seeking further information, such as the agent's father's condition or more background information normally available in the medical records.

Analytics Data

A total of 79 scenarios were captured and uploaded to the database during the workshops. In some instances, some scenarios may not have been captured because of participant error or early disconnection. In the 79 scenarios, a total of 416

XSL•FO
RenderX

"interactions" with the AI occurred. Of these 416 interactions, 96 had an empty STT input, which indicates nonrecognition of what the participant had said. This may have occurred because of a misalignment between when the participant pressed the button to start recording and when they started verbally responding to the agent.

Discussion

Principal Findings

The results of this study provide novel insights into the potential benefits of using verbal interactions with a VR-based conversational agent to supplement traditional verbal de-escalation training. It has been shown that immersive high-fidelity simulation can be beneficial over participant-to-participant role play due to a higher level of authenticity presented to enable the learning of verbal skills [28]. However, to date, there is little evidence of how such simulations can be translated into virtual settings with an elevated level of accuracy. This is principally due to the technology still being in relative infancy as well as the core functionality of speech recognition having identified challenges for more specialized use [29].

We conducted usability testing with feedback collated from the SUS form and free-text comments, as well as detailed observations taken throughout the study. The findings indicated the feasibility of the 8 factors identified to be key to the experience of users in VR training applications [12]. This study can provide a useful design direction for Code Black VR and future VR applications for clinical training. The feedback also highlighted several factors that impacted the experience, specifically within the Code Black VR, which were not part of the existing framework, motion sickness, perceived value, and privacy. We believe that these findings were identified because new factors included in this study were not observed in our previous work. The larger number of participants increased the likelihood of having participants who experienced certain elements of motion sickness, which is a well-documented side effect for some users [30]. We believe the "perceived value" factor arose in this study as it speaks to the highly specialized context of this application. Verbal de-escalation skills are perceived as a vital skill by and for clinicians; however, there is no real standardized way in which the content is delivered. Traditionally, these are skills established over years of service with significant complexity and nuance in their execution [3]. The factor of privacy was an unexpected but valuable outcome from this study. In previous studies, authors had conducted VR-based trials with small groups or individual interventions. The nature of this study and the logistics of completing it during the available time slots meant that larger groups were undertaking the experiences in closer proximity. This resulted in an innate awareness of the proximity of others and highlighted the need for this to be considered in future deployment strategies.

The average usability score of 59.5 placed Code Black VR's usability into the "poor" ranking on the usability scale. An SUS score below 68 indicates issues with design that require research and resolution [23]. This was not surprising, as the aim of this study was to test the feasibility of a newly developed prototype with significant scope for growth. The SUS tool helped the team identify areas of improvement to focus on in the next phase. One example was the need for improved orientation to the application; for example, through an interactive tutorial. We had acknowledged this need previously but underestimated its importance, given the perceived "natural verbal interactions" that the application was incorporating. This provides useful insight for future development and research in this area.

One surprising observation we made was the difference in perceived usability between the educator group (average score 50.6) and the clinician group (average score 61.6; Table 1). We suggest that this disparity arises from the different frames of reference of the more skilled and experienced clinicians and the perceived requirements of experience to train high-level verbal de-escalation skills. This was compared with the understanding of the requirements of the more junior group of clinicians in the study group. This finding aligns with our targeting of the application toward novice or intermediate entrants to the profession to develop basic levels of skills. The ultimate outcome is for high-level dialogue with a virtual agent; however, this requires significant further research and development of the AI system to reach that level of maturity, which is not achievable in the short term for these projects. Understanding the differences between the experience of the 2 groups provides a useful distinction for future studies on VR apps for clinical training and one that has not been addressed in the literature.

The analytics data allowed us to determine the number of completed scenarios and interactions and confirmed that the technology was functional. The empty STT inputs further reinforced the need to revise not only the way the user is orientated to the experience but also the way the interactions are captured within the device. When clinicians were engaged in dialogue, we witnessed several occasions where the clinicians "forgot" to use the push-to-talk function before responding to the agent. We are currently exploring the use of continual recording of user responses to minimize these dropouts; however, this brings different challenges, such as unintended responses, background noise, and misalignment with overlap occurring between the user and agent response cycles.

Challenges and Opportunities in Verbal De-escalation Training in VR

The thematic analysis of the participant responses highlights potential opportunities and limitations for using a combination of VR and conversational agents to address the challenges of developing verbal de-escalation skills. We predict a rapid growth in the application of VR to supplement clinical education in the future, as evidenced by the number of innovations occurring in our LHD. Despite our previous work identifying "user needs" in the design of VR-based applications [11,12], we identified several challenges in applying our knowledge in the context of the Code Black VR application, because this application features a conversational agent. The complexities of human language with variations in sentence structure, accents, and training are challenging to accommodate in an AI setting [31]. Therefore, while we are attempting to enhance user agency and allowing for advanced roles to align with the clinician's mental models,

XSL•FO
RenderX

the technical limitations faced by the current build undermine both usability and user experience. We address some of the ways in which these factors could be addressed in the *Code Black VR—Verbal De-escalation Skills Trainer: Next Steps* section.

Code Black VR—Verbal De-escalation Skills Trainer: Next Steps

This study identified several areas where Code Black VR can be improved to enhance the usability and user experience. The improvements and next steps are addressed through language recognition and generation, supplemental education, intra-application efficiencies, and broader application of code content.

Language Recognition and Generation

We note that a next step will involve identifying how to interpret, more accurately, the long-form answers that clinicians were providing within their dialogue. This could be achieved by improving the conversation structure by using a staged diatribe approach. This approach will connect user intents together, tagging intents within contextual sections and bringing a more structured approach to the dialogue. This will allow for greater context and improved accuracy in the response system. This approach also coincides with the implementation of a keyword search through the multiple intents to further increase interpretation accuracy.

We also suggest implementing a process in which the user response recording is always being recorded and a few seconds of recording spoken before the interaction button press is considered in the submission to the AI. This implementation will assist the experience in 2 important ways. First, it provides a solution for late button presses when the user responds before pressing the interaction button. Second, the DriftAI system will interpret what the user is saying and compare this to previous responses to increase the accuracy of the STT interpretation and subsequent response.

Another next step will be to improve the way DriftAI handles repetition. Currently, the same response was observed to be triggered multiple times in a row. Even with broader dialogue options, this occurrence is still a possibility. The developers are working on a way to reduce the repetition by tagging some responses as “unrepeatable” to ensure it is only matched once.

Following the completion of the study reported in this paper, we have also begun trialing a more sophisticated “sandbox” mode to provide an even larger training library to the DriftAI system. This allows for a more realistic and open dialogue with the agent; however, more development is required to control the AI, so it is aware of the context and stays “on track” to meet the desired learning outcomes.

Supplemental Education

The simulator, in its current state, requires clinicians to understand the basic verbal de-escalation skills to be successful. As such, we suggest implementing a module for novices to train them in these skills before using the simulator elements.

Intra-Application Efficiencies

In conjunction with the verbal de-escalation trainer, we are also deploying several other VR-based applications to support code black training for clinicians. These applications use other VR-based methodologies, such as 360° video of immersive clinical scenarios, auto stitching of the different views to allow the randomization of the 360° videos, and interactive hot spots to use newly adopted observation charts. The objectives of these applications are to support other vital code black skills, such as situational awareness, team planning, the detection of early signs of escalation, role allocation, and observations using the Behaviours of Concern observation chart. The goal is to embed all these modules within a single interconnected training package to better support the needs of the target clinicians. When this occurs, further research of the educational outcomes and effectiveness of the complete package will be conducted.

Broader Application of Core Content

We identified broader use cases for the underlying DriftAI technology that enables verbal interactions to occur. We will strive to use the findings of this study to adapt this verbal interaction approach to other clinical training situations using VR, flat-screen television, and mobile deployments.

Limitations

The breadth of the participant group in our study was indicative of emergency staff experience and skill sets in our LHD. However, the Code Black VR app is planned to be used in future iterations outside the ED setting. As such, further testing with users not from an ED background will be essential for further development and deployment.

In addition, the build of the application tested was still in the preliminary stages of development. As such, there were limitations on the available responses of the agent and the ability of the STT system to interpret some of the longer responses that clinicians provided to the agent. It was an intentional decision to conduct the study on such an early build of the application to explore the feasibility of the approach, but future studies on the developed application will be required to better understand the user experience and potential educational outcomes.

Another limitation of this study was in the method of limiting data collection to written forms and observations. A greater depth of understanding would likely have been established with the inclusion of user interviews during the data collection phase. The reason this was not included is related to the availability of participants who were enrolled in the study on the condition that the study be accommodated into part of their training program, which did not leave adequate time for the interview. We plan to implement user interviews in later studies on further developed iterations of the application; it was unfortunately not feasible for this study.

We acknowledge that the novel nature of VR can result in the “novelty effect” [32], which could account for some of the findings. In addition, further research and comparative studies should be conducted to explore the sustained adoption and knowledge retention of such novel approaches.

XSL•FO
RenderX

Conclusions

The implementation of effective verbal de-escalation training is essential to ensure the ongoing safety of both patients and health care providers. Simulation and face-to-face training are established approaches to deliver this type of training and increase staff confidence in these confronting situations. The challenge with the traditional approach of face-to-face simulation training is that it is asynchronous and resource intensive. The use of VR to supplement the existing training approaches is emerging as a feasible option. Our novel approach to using verbal interactions within VR was well received by clinicians. Our proposed framework with 11 factors could provide a much-needed direction for additional training assets to support clinicians faced with these challenging situations. The findings from this research contribute to the feasibility of our framework to support future research and the development of VR for clinical training. We have also contributed a set of requirements to guide the design of verbal interactions in the VR de-escalation training environment.

Conflicts of Interest

None declared.

References

1. Li Y, Li R, Qiu D, Xiao S. Prevalence of workplace physical violence against health care professionals by patients and visitors: a systematic review and meta-analysis. Int J Environ Res Public Health 2020 Jan 01;17(1):299 [FREE Full text] [doi: 10.3390/ijerph17010299] [Medline: 31906306]
2. McKnight S. De-Escalating Violence in Healthcare. Indianapolis, IN: Sigma Theta Tau Intl; 2019.
3. Davids J, Murphy M, Moore N, Wand T, Brown M. Exploring staff experiences: a case for redesigning the response to aggression and violence in the emergency department. Int Emerg Nurs 2021 Jul;57:101017. [doi: 10.1016/j.ienj.2021.101017] [Medline: 34174545]
4. Adams J. Assessing the effectiveness of clinical education to reduce the frequency and recurrence of workplace violence. Aus J Advanced Nursing 2017;34(3):6-15. [doi: 10.1097/01.nurse.0000282706.14770.f7]
5. Morphet J, Griffiths D, Beattie J, Velasquez Reyes D, Innes K. Prevention and management of occupational violence and aggression in healthcare: a scoping review. Collegian 2018 Dec;25(6):621-632. [doi: 10.1016/j.colegn.2018.04.003]
6. Aebersold M. Simulation-based learning: no longer a novelty in undergraduate education. Online J Issues Nurs 2018 Apr 03;23(2). [doi: 10.3912/ojin.vol23no02ppt39]
7. Bingham AL, Sen S, Finn LA, Cawley MJ. Retention of advanced cardiac life support knowledge and skills following high-fidelity mannequin simulation training. Am J Pharm Educ 2015 Mar 17;79(1):12 [FREE Full text] [doi: 10.5688/ajpe79112] [Medline: 25741028]
8. E‐learning for health professionals. Cochrane Database of Systematic Reviews. 2018. URL: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011736.pub2/full [accessed 2022-05-18]
9. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 08;9(3):e29080 [FREE Full text] [doi: 10.2196/29080] [Medline: 34255668]
10. Behind Angry Stan, the virtual reality simulator training nursing students to deal with conflict. ANMJ. URL: https://anmj.org.au/behind-angry-stan-the-virtual-reality-simulator-training-nursing-students-to-deal-with-conflict/ [accessed 2022-05-18]
11. Moore N, Yoo S, Ahmadpour N, Tommy R, Brown M, Poronnik P. ALS-SimVR: advanced life support virtual reality training application. In: Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology. 2019 Presented at: VRST '19: 25th ACM Symposium on Virtual Reality Software and Technology; Nov 12 - 15, 2019; Parramatta NSW Australia. [doi: 10.1145/3359996.3365051]
12. Moore N, Yoo S, Poronnik P, Brown M, Ahmadpour N. Exploring user needs in the development of a virtual reality-based advanced life support training platform: exploratory usability study. JMIR Serious Games 2020 Aug 07;8(3):e20797 [FREE Full text] [doi: 10.2196/20797] [Medline: 32763877]
13. King D, Tee S, Falconer L, Angell C, Holley D, Mills A. Virtual health education: scaling practice to transform student learning: using virtual reality learning environments in healthcare education to bridge the theory/practice gap and improve patient safety. Nurse Educ Today 2018 Dec;71:7-9. [doi: 10.1016/j.nedt.2018.08.002] [Medline: 30205259]
14. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
15. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
16. Immersive virtual reality and the developing child. In: Cognitive Development in Digital Contexts. Cambridge, Massachusetts, United States: Academic Press; 2017.

XSL•FO
RenderX

17. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001 Nov 1;25(11):1478-1483. [doi: 10.1007/s00268-001-0133-1] [Medline: 11760752]
18. Redefining home automation through voice recognition system. In: Emerging Technologies in Data Mining and Information Security. Singapore: Springer; 2021.
19. Soft skills training - reinvented. BodySwaps. URL: https://bodyswaps.co/ [accessed 2022-03-29]
20. Johnson M. Newcastle Uni taps biofeedback in VR conflict training. IT News. URL: https://www.itnews.com.au/news/newcastle-uni-taps-biofeedback-in-vr-conflict-training-534860 [accessed 2022-05-18]
21. Heckemann B, Hahn S, Halfens RJ, Richter D, Schols JM. Patient and visitor aggression in healthcare: a survey exploring organisational safety culture and team efficacy. J Nurs Manag 2019 Jul 29;27(5):1039-1046. [doi: 10.1111/jonm.12772] [Medline: 30888740]
22. Brooke J. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, Florida, United States: CRC Press; 1996.
23. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
24. Oculus Quest 2. Meta. URL: https://www.oculus.com/quest-2/ [accessed 2022-05-18]
25. Unreal Engine homepage. Unreal Engine. URL: https://www.unrealengine.com/en-US/ [accessed 2022-05-18]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Holton E, Swanson RA, Naquin SS. Andragogy in practice: clarifying the andragogical model of adult learning. Performance Improvement Q 2001;14(1):118-143. [doi: 10.1111/j.1937-8327.2001.tb00204.x]
28. Tufford L, Asakura K, Bogo M. Simulation versus role-play: perceptions of prepracticum BSW student. J Baccalaureate Social Work 2018;23(1):249-267. [doi: 10.18084/1084-7219.23.1.249]
29. Latif S, Rana R, Khalifa S, Jurdak R, Qadir J. Deep representation learning in speech processing: challenges, recent advances, and future trends. arXiv 2020.
30. Munafo J, Diedrick M, Stoffregen TA. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. Exp Brain Res 2017 Mar 3;235(3):889-901. [doi: 10.1007/s00221-016-4846-7] [Medline: 27915367]
31. Latif S, Qadir J, Qayyum A, Usama M, Younis S. Speech technology for healthcare: opportunities, challenges, and state of the art. IEEE Rev Biomed Eng 2021;14:342-356. [doi: 10.1109/rbme.2020.3006860]
32. Checa D, Miguel-Alonso I, Bustillo A. Immersive virtual-reality computer-assembly serious game to enhance autonomous learning. Virtual Real 2021 Dec 23:1-18 [FREE Full text] [doi: 10.1007/s10055-021-00607-1] [Medline: 34961808]

Abbreviations

AI: artificial Intelligence
ALS: advanced life support
ED: emergency department
HMD: head-mounted display
LHD: Local Health District
STT: speech to text
SUS: System Usability Scale
VR: virtual reality
WSLHD: Western Sydney Local Health District

Edited by N Zary; submitted 12.04.22; peer-reviewed by F Alvarez-Lopez, A Bustillo, D Lerner; comments to author 05.05.22; revised version received 17.05.22; accepted 12.06.22; published 06.07.22.

Please cite as:
Moore N, Ahmadpour N, Brown M, Poronnik P, Davids J
Designing Virtual Reality–Based Conversational Agents to Train Clinicians in Verbal De-escalation Skills: Exploratory Usability Study
JMIR Serious Games 2022;10(3):e38669
URL: https://games.jmir.org/2022/3/e38669
doi:10.2196/38669
PMID:35793129

XSL•FO
RenderX

©Nathan Moore, Naseem Ahmadpour, Martin Brown, Philip Poronnik, Jennifer Davids. Originally published in JMIR Serious Games (https://games.jmir.org), 06.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of the “Art Coloring” Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation

JuZhe Xi[1*], PhD; YuHan Gao[1*], MSc; Na Lyu[2], MSc; Zhuang She[1], MSc; XinYue Wang[3], BA; Xin-An Zhang[4], PhD; XiaoYu Yu[5], PhD; WeiDong Ji[6], MD; MengSheng Wei[7], MA; WeiHui Dai[7], PhD; Xuesheng Qian[7], PhD

[1]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

[2]Xinhua Hospital, Shanghai Jiao Tong University of Medicine, Shanghai, China

[3]College of Letters and Science, University of California, Berkeley, CA, United States

[4]Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai, China

[5]School of Management, Shanghai University, Shanghai, China

[6]Affiliated Mental Health Center (ECNU), Shanghai Changning Mental Health Center, Shanghai, China

[7]School of Management, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Xuesheng Qian, PhD
School of Management
Fudan University
Room 722, Siyuan Building
670 Guoshun Road
Shanghai, 200433
China
Phone: 86 13916208229
Email: qianxuesheng@fudan.edu.cn

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2022/3/e41253

Abstract

Background: COVID-19 has spread worldwide and generated tremendous stress on human beings. Unfortunately, it is often hard for distressed individuals to access mental health services under conditions of restricted movement or even lockdown.

Objective: The study first aims to develop an online digital intervention package based on a commercially released coloring game. The second aim is to test the effectiveness of difference intervention packages for players to increase subjective well-being (SWB) and reduce anxiety during the pandemic.

Methods: An evidence-based coloring intervention package was developed and uploaded to an online coloring game covering almost 1.5 million players worldwide in January 2021. Players worldwide participated to color either 4 rounds of images characterized by awe, pink, nature, and blue or 4 rounds of irrelevant images. Participants' SWB and anxiety and the perceived effectiveness of the game in reducing anxiety (subjective effectiveness [SE]) were assessed 1 week before the intervention (T1), after the participants completed pictures in each round (T2-T5), and after the intervention (T6). Independent 2-tailed *t* tests were conducted to examine the general intervention (GI) effect and the intervention effect of each round. Univariate analysis was used to examine whether these outcome variables were influenced by the number of rounds completed.

Results: In total, 1390 players worldwide responded and completed at least 1 assessment. Overall, the GI group showed a statistical significantly greater increase in SWB than the general control (GC) group (N=164, t162=3.59, Cohen d=0.59, 95% CI 0.36-1.24, *P*<.001). Compared to the control group, the best effectiveness of the intervention group was seen in the awe round, in which the increase in SWB was significant (N=171, t169=2.51, Cohen d=0.39, 95% CI 0.10-0.82, *P*=.01), and players who

XSL•FO
RenderX

colored all 4 pictures had nearly significant improvements in SWB (N=171, F4,170=2.34, partial ŋ2=0.053, *P*=.06) and a significant decrease in anxiety (N=171, F4,170=3.39, partial ŋ2=0.075, *P*=.01).

Conclusions: These data indicate the effectiveness of online psychological interventions, such as coloring games, for mental health in the specific period. They also show the feasibility of applying existing commercial games embedded with scientific psychological interventions that can fill the gap in mental crises and services for a wider group of people during the pandemic. The results would inspire innovations to prevent the psychological problems caused by public emergencies and encourage more games, especially the most popular ones, to take more positive action for the common crises of humankind.

(JMIR Serious Games 2022;10(3):e37026) doi:10.2196/37026

KEYWORDS

coloring game; online intervention; mental health; COVID-19 pandemic; gamification; game-based intervention; commercially released game

Introduction

COVID-19 has been spreading rapidly, with 237,890,892 confirmed cases and 4,852,328 deaths worldwide as of October 2021 [1]. The COVID-19 pandemic has threatened people's physical health as well as psychological health. Distress, depression, anxiety, posttraumatic stress symptoms, and insomnia are prevalent among people suffering from the pandemic [2,3]. Although emergency measures, such as social distancing and lockdown, helped prevent the spread of COVID-19, they also reduced people's access to social support and mental health services. Thus, accessing psychological aid became critical during the pandemic.

Numerous interventions have been developed to improve people's mental health during the COVID-19 crisis. These interventions typically use counseling services or self-guided programs for specific populations, including patients with COVID-19, health care workers, and people with mental illness [4,5]. However, the demerits of these interventions include limited popularization and a focus on narrow segments of people affected by the pandemic. Given the high prevalence of psychological distress (eg, anxiety, stress), new intervention models are needed to make mental health services accessible to the general population.

A promising solution is online self-guided interventions, which have the advantages of overcoming spatial barriers and being accessible to a broad population. For example, participants in Sweden who completed a 3-week online intervention reported significantly lower scores of worry [6]. However, the usefulness of such online interventions is largely restricted in experiments as most of them are not engaging enough. For the general population, it is often hard to adhere to such intervention programs due to limited interest in them.

We turn to digital games to address this issue. Digital games are entertaining in nature and have the potential to be designed into an attractive and effective intervention. For example, a recent study showed that virtual reality–based digital video games are effective in the treatment of arachnophobia [7]. This suggests that digital games have potential value as interventions to improve mental health. However, no game feature so far has been effective in improving mental health without degrading the game experience during the pandemic, especially commercially released games.

Coloring has been proved to reduce psychological distress and improve mood [8-10]. Coloring therapy is based on the idea that coloring can help people get away from their negative "inner dialogue" and negative thoughts and emotions [11]. This therapy has drawn wide research attention because coloring books have become popular among adults to relieve stress [12,13]. A study in Taiwan found that older adults who completed mandala coloring reported significantly lower anxiety levels and higher levels of calm, safety, and satisfaction than those in the control group [9]. In addition, online coloring games are currently popular (eg, Art Coloring & Color by Number, April Coloring, Lake Coloring), with a great number of fans worldwide. This alternative to digital coloring books has lower barriers to participation, more accessible operation, and more flexible cross-regional delivery. Although in the report, 1 limitation of physical coloring games is that unfinished coloring may cause additional psychological burdens [8], user-friendly online coloring can greatly simplify the coloring process and avoid it. Another limitation is that the improvement may be short term [8], but it also could be used as an accessible intervention for the general population during this particular period.

Four elements have been widely shown to be helpful in previous empirical studies [14-25]. First, coloring that elicits awe can reduce anxiety and strengthen happiness. Awe may transfer people's attention from themselves to a bigger picture and increase positive feelings, such as connectedness and humility, which benefit people, especially when facing adverse events [14,15]. Second, coloring that provides a sense of being in nature, such as viewing nature pictures or connecting with the natural environment, can improve mood and reduce anxiety [16,17]. Contact with nature may help people recover from stressful situations and improve attention concentration, subsequently enhancing mental health [18,19]. Third, the color pink is positively related to warmth, love, and nurturance, effectively reducing stress and violent behaviors within even minutes of exposure [20-23]. Fourth, the color blue, a cold color with a short wavelength, has calming and soothing effects and the function to benefit mood and reduce depression and anxiety [23-25]. The effectiveness of awe, pink, nature, and blue in reducing psychological distress suggests that coloring images characterized by these 4 elements online could be useful in reducing distress and improving psychological well-being during the COVID-19 pandemic.

XSL•FO
RenderX

In this study, we developed an online coloring package based on the commercially released Art Coloring app to address mental health issues during the pandemic. The advantage is that the coloring game can include a large number of players across regions, who can directly access the intervention and avoid the entry and learning threshold of a new game. We implemented the coloring package using 4 evidence-based image types (awe, pink, nature, and blue). We tested the effectiveness of the online coloring game in increasing subjective well-being (SWB) and reducing anxiety. We hypothesized that the digital gamification mental intervention would be effective in the pandemic; in addition, participants who colored specific types of images would show higher levels of improvement in SWB and lower levels of anxiety than participants who colored other images.

Methods

Ethical Considerations

The study was approved by East China Normal University's research ethics committee (HR151-2020).

Procedure and Participants

A commercially released app-based art drawing game called *Art Coloring - Coloring Book & Color by Number*, developed by Shiyi Network (Boke Technology Co., Ltd Company), was used as the intervention. The game has 1.5 million downloads, with over 80,000 daily active users (DAU) worldwide. Players worldwide can download the game on mobile phones or iPads from Google Play Store or Apple App Store and color line pictures following a series of numbered blocks to create colored pictures (Figure 1).

The participants in this study were existing players of the game. In addition to the basic coloring function, the game functions mainly involved in the study were "Daily Push" and "Daily News." Daily Push automatically sent all users 8 basic pictures daily, and Daily News published game notices nonscheduled.

In the study, Daily Push sent the 4 intervention coloring pictures mixed with 4 irrelevant pictures to players on schedule, following the group round order of awe, pink, nature, and blue (Figure 2); 1 group of pictures for 4 days was defined as 1 round of coloring. Every day, all players were free to choose and color the intervention images or irrelevant images (that do not have any of the 4 elements) and provide feedback through the Polls for Pages questionnaire embedded in Daily News. The procedure is shown in Figure 3. The questionnaire encouraged participants to leave email addresses and promised to draw 20 participants among all recorded emails to reward them with US $20 Amazon gift cards. The email addresses would be used as match codes, and cookies were placed on the players' devices to avoid duplicate submissions.

As shown in Figure 3, the study was conducted from January 29 to February 27, 2021. Every day for 16 days, all players were sent 1 group of coloring images. The participants' SWB and anxiety and the perceived effectiveness of the game in reducing anxiety (subjective effectiveness [SE]) of the participants were measured immediately after each round and right after the intervention. Self-report questionnaires were administered at baseline (T1), after each round of coloring was completed (T2-T5), and right after the intervention (T6). The questionnaire administrated after the intervention (T6) aimed to collect data from more participants, especially those who did not complete the last-round questionnaire. The methodology ensured that, in addition to questionnaires, all players' gaming experience would be the same as usual, which is the goal of this study—achieve an unnoticeable psychological enhancement when dealing with the psychological threat of a major crisis.

Two types of assessment were conducted to assess the intervention's effectiveness, both based on repeated measures. The general intervention (GI) effect was calculated based on the completed baseline and postintervention assessment of SWB, anxiety, and SE. The effect of each round of the intervention was calculated based on the assessment after each round.

Sociodemographic characteristics were collected at baseline, including the participants' gender, age, country, and average time spent on the game per day.

Outcome variables were SWB, anxiety, and SE. Each outcome was measured with 1 item to ease the participants' burden. SWB was measured using the Face Scale, a widely used single-item tool on which the participant rates their SWB by choosing 1 of 7 faces [26]. Anxiety was measured using the question "How have you been these 2 days?" Players rated their emotional state from 1 (very anxious) to 8 (very relaxed). The item was reverse-scored, with higher scores suggesting a higher level of anxiety. The SE was measured using the question "How effective do you think art coloring is in decreasing your anxiety level in general?" Players were asked to rate the item from 1 (no effect) to 8 (very effective).

XSL•FO
RenderX

Figure 1. “Art Coloring - Coloring Book & Color by Number” in Apple App Store.

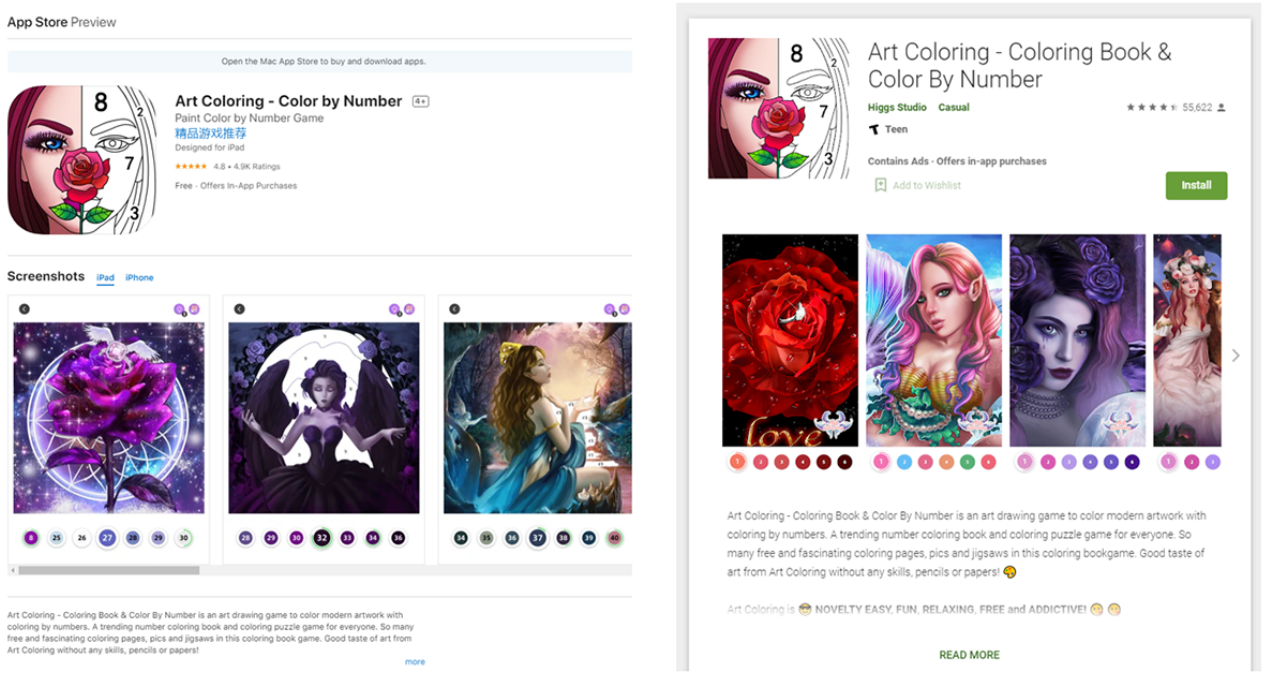
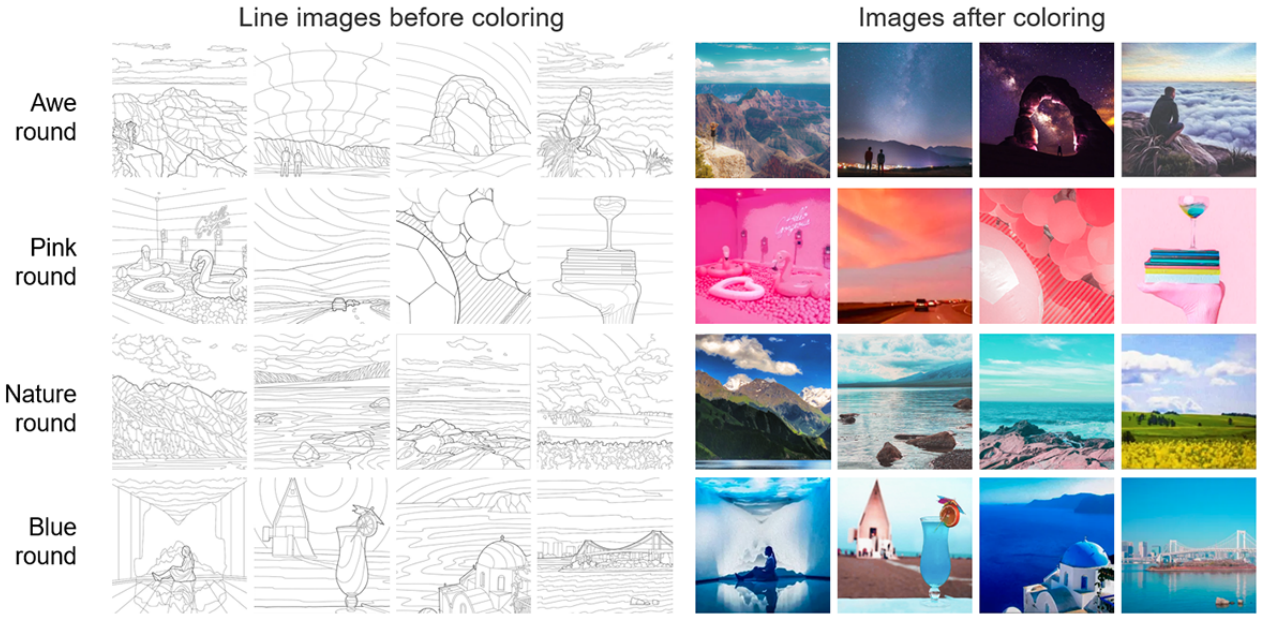


Figure 2. Coloring pictures for the awe, pink, nature, and blue rounds.



XSL•FO
RenderX

Figure 3. The study procedure.



Statistical Analysis

To control for baseline scores, residual postintervention scores were calculated. In addition, the intervention and control groups were compared to examine the intervention effect of each category of images. The number and group of players in each round varied depending on whether the player chose to color the image assigned for that round, to color unrelated images, or to not do any coloring. For those who only completed 1 postround questionnaire after the baseline assessment, residual scores were calculated as the difference between postround scores and baseline scores. Residual scores of those who completed the prior round were calculated as the difference between the current postround and prior postround scores. The total number of intervention package pictures that the participants colored was coded for analysis as follows: 0 (did not color any experimental pictures), 1 (colored 1-4 pictures), 2 (colored 5-8 pictures), 3 (colored 9-12 pictures), and 4 (colored 13-16 pictures).

Statistical analysis was conducted using SPSS Statistics 21.0 (IBM Corporation). Independent 2-tailed *t* tests were conducted to compare the intervention and control groups on the residual scores for the dependent variables (SWB, anxiety, and SE). Independent 2-tailed *t* tests and chi-square tests were conducted to investigate differences in independent variables between groups. Univariate analysis was performed to examine the association between the number of completed rounds and outcome variables. Covariates were entered to control for any baseline differences between the groups in the analyses.

Results

Participant Characteristics

In total, 1390 players from 31 countries ($N_{United\ States}$=1061, 76.3%; $N_{United\ Kingdom}$=124, 8.9%; N_{Canada}=51, 3.7%; N_{France}=33, 2.4%; $N_{Germany}$=32, 2.3%; $N_{Australia}$=23, 1.6%; N_{others}=66, 4.7%) responded to the study and completed at least 1 assessment, of which 612 (44%) completed the baseline assessment. Of these 612 participants with baseline data, 248 (41%) completed at least 1 postround assessment or the postintervention assessment, and these participants were kept in the final sample. The participants were divided into the intervention group (colored at least 1 experimental picture in that round) and the control group (did not complete any experimental pictures in that round); see Table 1. Of the 248 participants in the final sample, 164 (66%) completed the postintervention assessment. Of these 164 participants, 112 (68%) were in the GI group because they colored at least 1 of the 16 experimental pictures and 52 (32%) were in the general control (GC) group because they did not color any experimental pictures. The study trial profile is shown in Figure 4.

The majority (n=153, 93%) of the 164 participants were over 26 years old, 146 (89%) were female, and 124 (76%) spent over 30 minutes per day playing the online game at baseline (Table 1). Players aged 26-35 years (95% CI –2.76 to –0.41, *P*=.01), 46-55 years (95% CI –2.48 to –0.22, *P*=.02), and over 56 years (95% CI –2.45 to –0.26, *P*=.02) thought the art coloring was more effective in reducing anxiety than those under 25 years old. At baseline, no significant difference was found between the GI group and the GC group in SWB (t_{162}=–1.77, Cohen d=–0.29, 95% CI –0.84 to 0.05, *P*=.08), anxiety (t_{162}=0.59, Cohen d=0.10, 95% CI –0.44 to 0.82, *P*=.56), SE (t_{162}=1.32, Cohen d=0.22, 95% CI –0.19 to 0.93, *P*=.19), age (χ^2_4=1.89, *P*=.76), and average time spent (χ^2_3=3.00, *P*=.39).

Independent 2-tailed *t* tests were conducted to compare changes in outcome variables from baseline to postintervention in the GI and GC groups. Players' scores in the GI group showed a significantly larger increase in SWB compared to the GC group (N=164, t_{162}=3.59, Cohen d=0.59, 95% CI 0.36-1.24, *P*<.001). Though no significant difference was found in change in anxiety (N=164, t_{162}=–1.03, Cohen d=–0.17, 95% CI –1.11 to 0.35, *P*=.30) or change in SE (N=164, t_{162}=–0.53, Cohen d=–0.09, 95% CI –0.66 to 0.39, *P*=.60) between the 2 groups (Table 2).

Statistics in specific rounds, in the awe round, 171 (69%) of the 248 players completed the baseline and postround assessment. Of these, the 89 (52%) participants in the intervention group

XSL•FO
RenderX

showed a significantly higher increase in SWB than the 82 (48%) participants in the control group (N=171, t_{169}=2.51, Cohen d=0.39, 95% CI 0.10-0.82, *P*=.013). Changes in anxiety (N=171, t_{169}=–1.26, Cohen d=–0.19, 95% CI –0.99 to 0.22, *P*=.21) and SE (N=171, t_{169}=–0.68, Cohen d=–0.11, 95% CI –0.62 to 0.30, *P*=.50) were not significantly different between the 2 groups. When the average time spent and age were taken as covariates, these conclusions were not affected.

There was an unexpected result in the pink, nature, and blue rounds. In the pink round, 149 (60%) participants completed baseline and postround assessments. Of these, 84 (56%) participants in the intervention group reported a significantly greater decrease in SWB than those in the control group (N=149, t_{147}=–2.36, Cohen d=–0.40, 95% CI –0.62 to –0.05, *P*=.02). There was no significant difference in changes in anxiety (N=149, t_{147}=0.33, Cohen d=0.06, 95% CI –0.43 to 0.61, *P*=.74) or changes in SE (N=149, t_{147}=0.78, Cohen d=0.14, 95% CI –0.22 to 0.50, *P*=.44) between the 2 groups. In the nature round, 128 (52%) players completed baseline and postround assessments. Of these, 71 (55%) participants in the intervention group showed no significant difference between the 2 groups in changes in SWB (N=128, t_{126}=1.21, Cohen d=0.22, 95% CI –0.15 to 0.61, *P*=.23) and changes in anxiety (N=128, t_{126}=–1.06, Cohen d=–0.23, 95% CI –0.82 to 0.25, *P*=.29) but a significantly higher increase in changes in SE in the intervention group compared to the 57 (45%) participants in the control group (N=128, t_{126}=2.08, Cohen d=0.37, 95% CI 0.02-0.80, *P*=.04). In the blue round, 118 (48%) players completed baseline and postround assessments. Of these, 71 (60%) included in the intervention group showed no significant difference in changes in SWB (N=118, t_{116}=1.27, Cohen d=0.24, 95% CI –0.15 to 0.70, *P*=.21) , changes in anxiety (N=118, t_{116}=–1.43, Cohen d=–0.27, 95% CI –0.97 to 0.16, *P*=.16), or changes in SE (N=118, t_{116}=0.89, Cohen d=0.17, 95% CI –0.20 to 0.53, *P*=.38). It was verified that when the average time spent and age were taken as covariates, these conclusions were not affected.

As a further test of the intervention's effectiveness, we compared the participants in terms of how many rounds and how many pictures they completed. Among participants who completed both baseline and postintervention assessments (N=164), those who completed at least 1 round of intervention showed a significantly greater increase in SWB than those who did not undergo the intervention (112, 68%, vs 52, 34%, $F_{4,163}$=3.52, partial η^2=0.081, *P*=.01). Moreover, compared to players who completed irrelevant pictures, those who completed 1-12 experimental pictures showed a significant increase in SWB after the intervention (N=164, $F_{4,163}$=3.25, partial η^2=0.076, *P*=.01).

In the specific experimental rounds, analyses were conducted to compare participants based on the number of pictures they colored in each round. Results of univariate analysis showed that participants who colored either 1 or 4 awe pictures gained a nearly significant greater increase in SWB than those who did not color awe pictures (N=171, $F_{4,170}$=2.34, partial η^2=0.053, *P*=.06). Those who completed 4 awe pictures had a significantly larger decrease in anxiety than those who completed one or did not color awe pictures (N=171, $F_{4,170}$=3.39, partial η^2=0.075, *P*=.01). No association was found between the number of colored-picture numbers and outcome variables in the pink, nature, or blue round (Table 3).

XSL•FO
RenderX

Table 1. Baseline demographic characteristics.

Characteristic	GI[a] group (N=112), n (%)	GC[b] group (N=52), n (%)	Awe round (N=171), n (%)		Pink round (N=149), n (%)		Nature round (N=128), n (%)		Blue round (N=118), n (%)	
			Group 1[c] (N=89)	Group 2[d] (N=82)	Group 1 (N=84)	Group 2 (N=65)	Group 1 (N=71)	Group 2 (N=57)	Group 1 (N=71)	Group 2 (N=47)
Gender										
Male	10 (9)	8 (15)	7 (8)	9 (11)	10 (12)	6 (9)	5 (7)	7 (12)	9 (13)	6 (13)
Female	102 (91)	44 (85)	82 (92)	73 (89)	74 (88)	59 (91)	66 (93)	50 (88)	62 (87)	41 (87)
Age (years)										
≤25	9 (8)	2 (4)	8 (9)	4 (4.9)	4 (5)	4 (6)	5 (7)	3 (5)	4 (6)	2 (4)
26-35	18 (16)	7 (13)	18 (20)	11 (13)	11 (13)	6 (9)	12 (17)	5 (9)	12 (17)	5 (11)
36-45	34 (30)	15 (29)	25 (28)	24 (29)	32 (38)	18 (28)	23 (32)	14 (25)	21 (30)	9 (19)
46-55	23 (21)	11 (21)	14 (16)	27 (33)	23 (27)	17 (26)	12 (17)	17 (30)	17 (24)	16 (34)
≥56	28 (25)	17 (33)	24 (27)	16 (20)	14 (17)	20 (31)	19 (27)	18 (32)	17 (24)	15 (32)
Average time spent (minutes)										
≤5	0 (0)	3 (6)	3 (3)	0 (0)	2 (2)	1 (2)	1 (1)	0 (0)	1 (2)	1 (2)
5-15	5 (4)	5 (10)	6 (7)	6 (7)	4 (5)	5 (8)	3 (4)	5 (9)	3 (4)	2 (4)
16-30	19 (17)	8 (15)	19 (21)	16 (20)	14 (17)	10 (15)	12 (17)	11 (19)	13 (18)	6 (13)
≥31	85 (76)	39 (75)	61 (69)	60 (73)	64 (76)	49 (75)	55 (78)	41 (72)	54 (76)	38 (81)

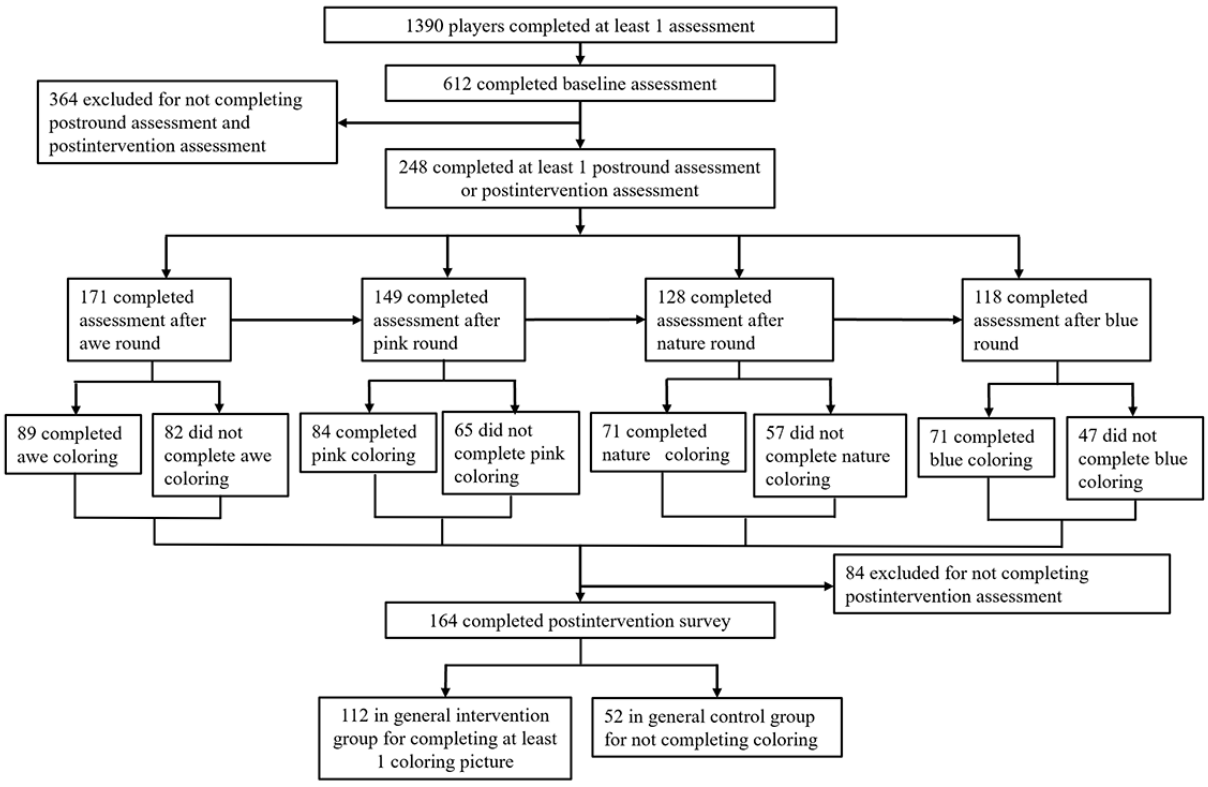
[a]GI: general intervention.

[b]GC: general control.

[c]Group 1: intervention group.

[d]Group2: control group.

Figure 4. Trial profile.



XSL•FO
RenderX

Table 2. Independent 2-tailed *t* tests between residual scores of the intervention and control groups (N=164).

Outcome variable	Intervention group, mean (SD)	Control group, mean (SD)	*t* test (*df*)	*P* value	Cohen d
Total (N=164)					
SWB[a]	0.24 (1.22)	–0.56 (1.53)	3.59 (162)	<.001	0.59
Anxiety	–0.82 (2.22)	–0.44 (2.14)	–1.03 (162)	.30	–0.17
SE[b]	–0.06 (1.49)	0.08 (1.77)	–0.53 (162)	.60	–0.09
AWE round (N=171)					
SWB	0.20 (1.29)	–0.26 (1.08)	2.51 (169)	.01	0.39
Anxiety	–0.48 (2.07)	–0.10 (1.91)	–1.26 (169)	.21	–0.19
SE	–0.15 (1.37)	0.01 (1.67)	–0.68 (169)	.50	–0.11
Pink round (N=149)					
SWB	–0.06 (0.72)	0.28 (1.02)	–2.36 (147)	.02	–0.40
Anxiety	–0.40 (1.65)	–0.49 (1.53)	0.33 (147)	.74	0.06
SE	0.10 (1.29)	–0.05 (0.78)	0.78 (147)	.44	0.14
Nature round (N=128)					
SWB	–0.08 (1.09)	–0.32 (1.05)	1.21 (126)	.23	0.22
Anxiety	–0.03 (1.26)	0.32 (1.79)	–1.06 (126)	.29	–0.23
SE	0.20 (1.19)	–0.21 (0.98)	2.08 (126)	.04	0.37
Blue round (N=118)					
SWB	0.46 (0.92)	0.19 (1.41)	1.27 (116)	.21	0.24
Anxiety	–0.66 (1.43)	–0.26 (1.62)	–1.43 (116)	.16	–0.27
SE	0.06 (1.07)	–0.11 (0.81)	0.89 (116)	.38	0.17

[a]SWB: subjective well-being.

[b]SE: subjective effectiveness (of art coloring in reducing anxiety).

XSL•FO
RenderX

Table 3. Univariate analysis between colored-picture numbers and outcome variables for each round.

Outcome variable	Colored-picture numbers in each round, mean (SD)					*F* test (*df*)	*P* value
	0	1	2	3	4		
Awe round (N=171)							
SWB[a]	–0.26 (1.08)	0.17 (1.34)	–0.23 (1.09)	0.27 (1.19)	0.48 (1.34)	2.34 (4,170)	.06
Anxiety	–0.10 (1.91)	0.17 (1.85)	–0.62 (2.06)	–0.55 (0.82)	–1.57 (2.48)	3.39 (4,170)	.01
SE[b]	0.01 (1.67)	–0.29 (1.20)	–0.15 (0.90)	–0.36 (1.03)	0.22 (1.93)	0.59 (4,170)	.67
Pink round (N=149)							
SWB	0.28 (1.02)	–0.08 (0.68)	0.07 (0.83)	–0.22 (0.83)	0.00 (0.71)	1.55 (4,148)	.19
Anxiety	–0.49(1.53)	–0.38 (1.77)	–0.36 (1.50)	0.33 (1.12)	–1.08 (1.55)	1.10 (4,148)	.36
SE	–0.05 (0.78)	0.23 (1.40)	–0.07 (1.33)	–0.11 (1.05)	–0.08 (0.95)	0.57 (4,148)	.69
Nature round (N=128)							
SWB	–0.32 (1.05)	0.09 (1.04)	0.00 (1.00)	–0.20 (0.45)	–0.31 (1.26)	0.88 (4,127)	.48
Anxiety	0.32 (1.79)	0.11 (1.21)	–0.60 (0.89)	–0.20 (1.10)	0.08 (1.44)	0.55 (4,127)	.70
SE	–0.21 (0.98)	0.26 (1.36)	0.80 (0.45)	–0.20 (0.45)	0.08 (1.13)	1.72 (4,127)	.15
Blue round (N=118)							
SWB	0.19(1.41)	0.33(0.84)	0.36(0.67)	1.17(1.60)	0.50(0.88)	1.10 (4,117)	.36
Anxiety	–0.26(1.62)	–0.53(1.47)	–0.38(1.09)	–0.88(1.86)	–0.17(0.96)	0.76 (4,117)	.56
SE	0.09(0.71)	–0.00(0.88)	0.38(1.02)	–0.32(1.46)	–0.08(0.58)	1.03 (4,117)	.39

[a]SWB: subjective well-being.

[b]SE: subjective effectiveness (of art coloring in reducing anxiety).

Discussion

Principal Findings

This study developed an online gamification intervention based on an existing global game to reduce anxiety and promote SWB during the COVID-19 pandemic. The online gamified intervention based on the game form of coloring considered as a relaxing and natural activity can prompt participants' mental state. The intervention packaged 4 types of pictures that have proved effective in previous research. We implemented a study to verify the effectiveness of both the online game–based method and intervention packages.

Taken together, participants in the GI group reported a significant increase in SWB compared to the GC group, even if they may be imperceptible. Previous studies have demonstrated that the repetition of coloring may produce a calming, almost trance-like effect [27], which could evoke participants' inner power and help them get out of negative thoughts and moods. The report showed visual art making is a relaxing experience for some [28]. However, the intervention effect of coloring was not significant in reducing anxiety or even decrease SWB in the GC group. Considering the increasing trend of the COVID-19 pandemic in the world, long-term and ongoing exposure to stressors related to COVID-19 may explain this finding. Coloring the 4 types of pictures and coloring itself that could reduce anxiety in peaceful times might successfully counteract the effect of stressors and prevent anxiety from increasing, rather than not being effective enough to reduce anxiety during an epidemic. The findings implied that some interventions proven effective in reducing psychological distress in prepandemic times might maintain the level of distress during the COVID-19 pandemic. Therefore, it should be noted that the study results of the online gamified coloring method for improving SWB and reducing anxiety may be underestimated during the pandemic.

In the specific intervention package, the awe-inspiring scenes were effective in increasing SWB. It is also noteworthy that those who completed all 4 awe pictures reported a significantly greater increase in SWB and a reduction in anxiety compared to the control group. It is possible that awe-inspiring scenes helped people to stay in the present moment, providing a sense of time expansion and creating positivity and satisfaction in their lives [15]. Another possible reason was that the awe experience distracted people from struggles in life, such as the chaos caused by COVID-19, and directed their attention to the light side of life [29]. People undergoing the outbreak of COVID-19, especially those going through lockdown, faced many uncertainties that could bring psychological distress [2]. Research has demonstrated that awe effectively mitigates anxiety in a stressful waiting period [14]. Therefore, the awe experience might reduce uncertainties by directing players' attention to a "big picture," improving the feeling of connectedness and promoting SWB during the stressful lockdown period.

Nevertheless, images that featured a lot of pink or blue or nature did not yield expected results, though they had positive reports. The approaches in peaceful times may not be effective enough to improve the mental state during an epidemic. Specifically, players who completed pink pictures reported a significant

XSL•FO
RenderX

decrease in SWB than those in the control group. This was probably because too much pink color was used in the pink round. Although pink is shown to have a tranquilizing and calming effect on negative emotions [20,22], too much pink has been found to make people physically drained and emasculated [30], which in turn could reduce their SWB. This finding suggests that pictures with too much pink should be excluded from coloring materials. Those with less pink could also be examined for the intervention effect during the COVID-19 pandemic. Similarly, blue might be perceived as either calm and soothing or as cold and unfriendly. These materials of dual character might not be suitable for extreme situations of the pandemic.

Interestingly, in the nature round, the participants' perceived effectiveness of the picture coloring did not reflect in the measure of SWB or anxiety. The participants might be influenced by the consensus on the approved effectiveness of nature pictures. Considering the challenging period all people were undergoing, coloring pictures of nature may release a positive signal for players to face difficult situations, which was also valuable during the pandemic. Another reason was likely to be rooted in the length of exposure to nature pictures. Previous studies have demonstrated that participants reported the most significant increase in self-esteem and mood after 5 minutes of exposure to nature pictures [31]. However, the majority of players in the nature round spent more than 5 minutes playing, which might lead to a reduction in the intervention effect of nature coloring. A simplified intervention using nature pictures that can be colored in around 5 minutes should be tested in future studies to examine the specific intervention effect of nature coloring. The third reason may be during the pandemic, people may gradually get used to the long-term stress caused by COVID-19 and tend to maintain their present anxiety levels. Future research is needed to examine the intervention effect of coloring on participants' confidence in dealing with difficult situations.

Strengths

Our research has several strengths. First, this study is the first to investigate the effect of an online coloring intervention on psychological well-being during the COVID-19 pandemic. This may give us a close view of how a coloring intervention works in a natural condition when people are undergoing a stressful period. Second, we demonstrated that the online digital coloring package improved people's SWB during this period and in certain circumstances may have reduced anxiety. Third, compared to conventional interventions, a game-based alternative may more easily evoke people's interest. The online format, with its accessibility, entertainment potential, and acceptability for people with barriers to seeking help, can be a practical approach to helping the general population during the pandemic. Fourth, the positive results and novel intervention approaches can stimulate more research of color and picture types as psychological intervention packages that can be delivered online so as to reduce the anxiety and other psychological distress in the general population during the pandemic.

Limitations

This study has its limitations. First, to maintain the user experience of a commercially released game, we abandoned the plan of forcing all players (about 100,000 people per day) to participate in the study. Instead, we tried to keep the game experience as the usual, which was the goal of this study—to provide an unnoticeable psychological enhancement when dealing with the psychological threat of a major crisis. However, the cost was that we could not give full play to the advantages of the millions of players of the popular online game.

Second, to minimize the burden on participants and to get a better response rate, each outcome was assessed using 1-item measures. These measures were selected to shorten the length of the questionnaire in order to reduce the players' burden and enlarge the size of the sample. The 1-item SWB question was assessed using the 1-item Face Scale, which has good psychometric properties. However, it is undeniable that large-scale standard scales can obtain more information.

Third, the gender distribution of the players in the coloring game was skewed. The players in the coloring game were predominantly female. Therefore, the research findings need further validation regarding the prevalence of the effects in both genders. However, the study is timely as to the immediate value of psychological interventions during the pandemic, considering that women are more prone to anxiety [32] and that screen time is negatively correlated with mental health [33].

Fourth, from the perspective of statistical calculation, the intervention effect of each round may have been underestimated, as the residual scores were calculated as the difference between each round and the one before it. The positive effects of one round might carry over into the next round, leading to underestimation of the intervention effect in the latter round.

Fifth, it is noteworthy that we considered the coloring intervention may have a significant long-term effect rather than an immediate effect on anxiety. However, online psychological intervention involves a complex superposition of psychological and cognitive mechanism, which cannot be fully demonstrated by separate experiments. Therefore, more game-based interventions and other coloring materials deserve further study, and long-term effects deserve continued attention.

Conclusion

This study developed an entertaining, accessible, cross-regional, and effective online gamification intervention to promote SWB in a community sample of adults during the COVID-19 pandemic. The game-based intervention prompted coloring as a relaxing activity. Participants colored images that included elements that have been shown to be relaxing in previous research: an awe-inspiring scene, a nature scene, and scenes dominated by pink or blue. In general, the intervention group showed a significantly greater increase in SWB compared to the control group, and awe images proved an advantage in the online coloring approach.

Finally, the study adopted an existing released online game with high global activity as the vehicle, which eliminated the difficulties of communication and learning of new tools and,

XSL•FO
RenderX

most importantly, showed the feasibility of applying existing commercial games embedded with scientific psychological interventions that can fill the gap in mental crises and services for a wider group of people during the pandemic. The result would inspire innovations to prevent the psychological problems caused by public emergencies and encourage more games, especially the most popular ones, to take more positive action for the common crises of humankind.

Acknowledgments

We would like to thank Boke Technology Co., Ltd, for game development and platform support. Yanan Peng and Shijia Li of the School of Psychology and Cognitive Science, East China Normal University, and Bin Wang of the School of Management, Shanghai University, also contributed to this paper.

This work was supported by the National Natural Science Foundation of China (Nos. 71971066, 71972126); the National Social Science Foundation of China (No. 18BSH129); the Major Project of Philosophy and Social Science Research, Ministry of Education of China (No. 19JZD010); the Shanghai Philosophy and Social Science Planning Project (No. 2017BSH004); the Shanghai Pujiang Program of China (No. 12PJ037); the Research Project of Shanghai Science and Technology Commission (No. 20DZ2260300); and the Innovation Program of Shanghai Municipal Education Commission (No. 2019-01-07-00-09-E00078).

Authors' Contributions

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had the final responsibility for the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. Gardner L. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). URL: https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6 [accessed 2021-10-11]
2. Krishnamoorthy Y, Nagarajan R, Saya GK, Menon V. Prevalence of psychological morbidities among general population, healthcare workers and COVID-19 patients amidst the COVID-19 pandemic: a systematic review and meta-analysis. Psychiatry Res 2020 Nov;293:113382 [FREE Full text] [doi: 10.1016/j.psychres.2020.113382] [Medline: 32829073]
3. van Agteren J, Bartholomaeus J, Fassnacht DB, Iasiello M, Ali K, Lo L, et al. Using internet-based psychological measurement to capture the deteriorating community mental health profile during COVID-19: observational study. JMIR Mental Health 2020 Jun 11;7(6):e20696 [FREE Full text] [doi: 10.2196/20696] [Medline: 32490845]
4. Zhang C, Yang L, Liu S, Xu Y, Zheng H, Zhang B. One-week self-guided internet cognitive behavioral treatments for insomnia in adults with situational insomnia during the COVID-19 outbreak. Front Neurosci 2020;14:622749 [FREE Full text] [doi: 10.3389/fnins.2020.622749] [Medline: 33551732]
5. Cheng P, Xia G, Pang P, Wu B, Jiang W, Li Y, et al. COVID-19 epidemic peer support and crisis intervention via social media. Community Ment Health J 2020 Jul;56(5):786-792 [FREE Full text] [doi: 10.1007/s10597-020-00624-5] [Medline: 32378126]
6. Wahlund T, Mataix-Cols D, Olofsdotter Lauri K, de Schipper E, Ljótsson B, Aspvall K, et al. Brief online cognitive behavioural intervention for dysfunctional worry related to the COVID-19 pandemic: a randomised controlled trial. Psychother Psychosom 2021 Nov 19;90(3):191-199 [FREE Full text] [doi: 10.1159/000512843] [Medline: 33212440]
7. Bouchard S, Côté S, St-Jacques J, Robillard G, Renaud P. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. Technol Health Care 2006;14(1):19-27. [Medline: 16556961]
8. Turturro N, Drake JE. Does coloring reduce anxiety? Comparing the psychological and psychophysiological benefits of coloring versus drawing. Empir Stud Arts 2020 May 06;40(1):3-20. [doi: 10.1177/0276237420923290]
9. Koo M, Chen H, Yeh Y. Coloring activities for anxiety reduction and mood improvement in Taiwanese community-dwelling older adults: a randomized controlled study. Evid Based Complement Alternat Med 2020 Jan 22;2020:6964737-69647376 [FREE Full text] [doi: 10.1155/2020/6964737] [Medline: 32063986]
10. Rajendran N, Mitra TP, Shahrestani S, Coggins A. Randomized controlled trial of adult therapeutic coloring for the management of significant anxiety in the emergency department. Acad Emerg Med 2020 Feb;27(2):92-99 [FREE Full text] [doi: 10.1111/acem.13838] [Medline: 31957143]
11. Curry NA, Kasser T. Can coloring mandalas reduce anxiety? Art Therapy 2005 Jan;22(2):81-85. [doi: 10.1080/07421656.2005.10129441]
12. Ashlock LE, Miller-Perrin C, Krumrei-Mancuso E. The effectiveness of structured coloring activities for anxiety reduction. Art Therapy 2019 Feb 07;35(4):195-201. [doi: 10.1080/07421656.2018.1540823]

XSL•FO
RenderX

13. Bender RG. 5 Unique Ways to Handle Stress That Actually Work. URL: https://www.businessinsider.com/5-unique-ways-to-handle-stress-that-actually-work-2015-8?IR=T [accessed 2022-06-20]
14. Rankin K, Andrews SE, Sweeny K. Awe-full uncertainty: easing discomfort during waiting periods. J Posit Psychol 2019 May 15;15(3):338-347. [doi: 10.1080/17439760.2019.1615106]
15. Rudd M, Vohs KD, Aaker J. Awe expands people's perception of time, alters decision making, and enhances well-being. Psychol Sci 2012 Oct 01;23(10):1130-1136. [doi: 10.1177/0956797612438731] [Medline: 22886132]
16. Vincent E, Battisto D, Grimes L, McCubbin J. The effects of nature images on pain in a simulated hospital patient room. Heal Environ Res Des J 2010;3(3):42-55. [doi: 10.1177/193758671000300306] [Medline: 21165860]
17. Yamashita R, Chen C, Matsubara T, Hagiwara K, Inamura M, Aga K, et al. The mood-improving effect of viewing images of nature and its neural substrate. Int J Environ Res Public Health 2021 May 20;18(10):5500 [FREE Full text] [doi: 10.3390/ijerph18105500] [Medline: 34065588]
18. Kaplan S. The restorative benefits of nature: toward an integrative framework. J Environ Psychol 1995 Sep;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
19. Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M. Stress recovery during exposure to natural and urban environments. J Environ Psychol 1991 Sep;11(3):201-230. [doi: 10.1016/s0272-4944(05)80184-7]
20. Pellegrini RJ, Schauss AG, Miller ME. Room color and aggression in a criminal detention holding cell: a test of the "tranquilizing pink" hypothesis. J Orthomol Psychiatry 1981;10(3):174-181.
21. Hamid PN, Newport AG. Effect of colour on physical strength and mood in children. Percept Mot Skills 1989 Aug 31;69(1):179-185. [doi: 10.2466/pms.1989.69.1.179] [Medline: 2780176]
22. Pakseresht M, Hemmatipour A, Gilavand A, Zarea K, Poursangbor T, Sakei-malehi A. The effect of nurses' uniform color on situational anxiety in the school age inpatients children. J Med Dent Sci 2019;7(1):114-120.
23. Lubos LC. The role of colors in stress reduction. Liceo J High Educ Res 2008 Dec 30;5(2):95-103. [doi: 10.7828/ljher.v5i2.39]
24. Grote V, Kelz C, Goswami N, Stossier H, Tafeit E, Moser M. Cardio-autonomic control and wellbeing due to oscillating color light exposure. Physiol Behav 2013 Apr 10;114-115:55-64. [doi: 10.1016/j.physbeh.2013.03.007] [Medline: 23506685]
25. Wahl S, Engelhardt M, Schaupp P, Lappe C, Ivanov IV. The inner clock: blue light sets the human rhythm. J Biophotonics 2019 Dec;12(12):e201900102 [FREE Full text] [doi: 10.1002/jbio.201900102] [Medline: 31433569]
26. Andrew FM, Withey SB. Social Indicators of Well-Being. New York; London: Plenum; 1976:696-717.
27. Sandmire DA, Gorham SR, Rankin NE, Grimm DR. The influence of art making on anxiety: a pilot study. Art Therapy 2012 Jun;29(2):68-73. [doi: 10.1080/07421656.2012.683748]
28. Kaimal G, Ray K, Muniz J. Reduction of cortisol levels and participants' responses following art making. Art Therapy (Alex) 2016 Apr 02;33(2):74-80 [FREE Full text] [doi: 10.1080/07421656.2016.1166832] [Medline: 27695158]
29. Büssing A, Wirth AG, Reiser F, Zahn A, Humbroich K, Gerbershagen K, et al. Experience of gratitude, awe and beauty in life among patients with multiple sclerosis and psychiatric disorders. Health Qual Life Outcomes 2014 Apr 30;12:63 [FREE Full text] [doi: 10.1186/1477-7525-12-63] [Medline: 24779860]
30. Schauss AG. Tranquilizing effect of color reduces aggressive behavior and potential violence. J Orthomol Med 1979 Dec 01;8(4):218-221 [FREE Full text] [doi: 10.1136/jnnp.42.12.1163] [Medline: 533856]
31. Brown DK, Barton JL, Gladwell VF. Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress. Environ Sci Technol 2013 Jun 04;47(11):5562-5569 [FREE Full text] [doi: 10.1021/es305019p] [Medline: 23590163]
32. Maaravi Y, Heller B. Not all worries were created equal: the case of COVID-19 anxiety. Public Health 2020 Aug;185:243-245 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.032] [Medline: 32688099]
33. Riehm KE, Holingue C, Kalb LG, Bennett D, Kapteyn A, Jiang Q, et al. Associations between media exposure and mental distress among U.S. adults at the beginning of the COVID-19 pandemic. Am J Prev Med 2020 Nov;59(5):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2020.06.008] [Medline: 33011008]

Abbreviations

GC: general control
GI: general intervention
SE: subjective effectiveness
SWB: subjective well-being

XSL•FO
RenderX

Edited by N Zary; submitted 03.02.22; peer-reviewed by H van Marwijk, A Teles, JW Tan, M Brown; comments to author 25.03.22; revised version received 24.04.22; accepted 13.05.22; published 08.07.22.

Please cite as:
Xi J, Gao Y, Lyu N, She Z, Wang X, Zhang XA, Yu X, Ji W, Wei M, Dai W, Qian X
Effect of the "Art Coloring" Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation
JMIR Serious Games 2022;10(3):e37026
URL: https://games.jmir.org/2022/3/e37026
doi:10.2196/37026
PMID:35575761

©JuZhe Xi, YuHan Gao, Na Lyu, Zhuang She, XinYue Wang, Xin-An Zhang, XiaoYu Yu, WeiDong Ji, MengSheng Wei, WeiHui Dai, Xuesheng Qian. Originally published in JMIR Serious Games (https://games.jmir.org), 08.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Serious Games for Improving Executive Functions Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis

Alaa Abd-alrazaq[1,2], PhD; Dari Alhuwail[3,4], PhD; Arfan Ahmed[1], PhD; Mowafa Househ[2], PhD

[1]AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

[2]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[3]Information Science Department, College of Life Sciences, Kuwait University, Kuwait, Kuwait

[4]Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
P.O. Box 34110
Doha Al Luqta St, Ar-Rayyan
Doha
Qatar
Phone: 974 55708546
Email: mhouseh@hbku.edu.qa

Abstract

Background: Executive functions are one of the known cognitive abilities that decline with age. They are the high-order cognitive processes that enable an individual to concentrate, plan, and take action. Serious games, which are games developed for specific purposes other than entertainment, could play a positive role in improving executive functions. Several systematic reviews have pooled the evidence about the effectiveness of serious games in improving executive functions; however, they are limited by some weaknesses.

Objective: This study aims to investigate the effectiveness of serious games for improving executive functions among older adults with cognitive impairment.

Methods: A systematic review of randomized controlled trials (RCTs) was conducted. To retrieve relevant studies, 8 electronic databases were searched. Further, reference lists of the included studies and relevant reviews were screened, and we checked studies that cited our included studies. Two reviewers independently checked the eligibility of the studies, extracted data from the included studies, assessed the risk of bias, and appraised the quality of the evidence. We used a narrative and statistical approach, as appropriate, to synthesize results of the included studies.

Results: Of 548 publications identified, 16 RCTs were eventually included in this review. Of the 16 studies, 14 studies were included in 6 meta-analyses. Our meta-analyses showed that serious games are as effective as no or passive interventions at improving executive functions (*P*=.29). Surprisingly, conventional exercises were more effective than serious games at improving executive functions (*P*=.03). Our subgroup analysis showed that both types of serious games (cognitive training games, *P*=.08; exergames, *P*=.16) are as effective as conventional exercises at improving executive functions. No difference was found between adaptive serious games and nonadaptive serious games for improving executive functions (*P*=.59).

Conclusions: Serious games are not superior to no or passive interventions and conventional exercises at improving executive functions among older adults with cognitive impairment. However, our findings remain inconclusive due to the low quality of the evidence, the small sample size in most included studies, and the paucity of studies included in the meta-analyses. Accordingly, until more robust evidence is available, serious games should not be offered by health care providers nor used by patients for improving executive functions among older adults with cognitive impairment. Further reviews are needed to assess the long-term effect of serious games on specific executive functions or other cognitive abilities among people from different age groups with or without cognitive impairment.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42021272757; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=272757

(JMIR Serious Games 2022;10(3):e36123) doi:10.2196/36123

KEYWORDS

serious games; cognitive training; exergames; executive functions; mild cognitive impairment; Alzheimer disease; dementia; systematic reviews; meta-analysis

Introduction

Background

Globally, the older adult population is rapidly increasing at unprecedented rates. By the year 2050, approximately 2 billion people are expected to live to over 65 years old [1]. As people age and live longer, it is unclear if their additional years of living are enjoyed in good health [2]. Generally, as people grow older, their risk of experiencing cognitive impairment increases [3]. After the age of 70 years, older adults will, unfortunately, experience physical and mental multimorbidities [4]; they will require special care and attention because of the emergence of multiple progressive health complications, including declining mental and cognitive functions, noncommunicable diseases (eg, diabetes), vision impairments, hearing loss, and physical ailments [5]. Additionally, aging is often accompanied by various social problems, including economic or financial insecurity, isolation, and loneliness [6].

For older adults, the World Health Organization (WHO) estimates that nearly 7% of the total disability-adjusted life years (DALYs) are attributed to mental and neurological disorders [7]. Among the top culprits causing the progressive decline in cognitive functions and abilities is mild cognitive impairment (MCI), which in turn increases the risk for developing dementia and Alzheimer disease (AD) [8]. In the United States alone, it is estimated that, by 2050, approximately 13.8 million older adults will have AD-related dementia [9]. Economically, European countries estimated the toll of AD alone at €232 billion in 2015, and it is expected that this cost will double by 2040 [10]. The toll brought by the declining mental and cognitive functions of older adults places a large financial burden on public health. The stress brought by the declining mental and cognitive functions of older adults is further exacerbated by effects on the older adult's family members, caregivers, and society. Therefore, the WHO recommended that the prevention of mental and cognitive decline is to be ranked as a global mental health priority [11].

One of the cognitive abilities that decline by age is executive function. Executive functions are essential for flexible, adaptive, and goal-oriented behavior [12]. Executive functions can be referred to as the high-order cognitive processes that enable an individual to concentrate, plan, and take actions [12]. Working memory, flexible thinking, and self-control are among the mental qualities that comprise executive function. Every day, we employ these abilities to learn, work, and govern our lives. Executive function issues can make it difficult to focus, follow directions, and manage emotions, to name a few things. Executive functions can generally be grouped into 3 core processes: (1) inhibiting predominant responses and controlling attention; (2) switching between tasks and cognitive flexibility; and (3) updating, retaining, and processing information [13,14]. Research suggests several ways to improve executive functions, including both pharmacological and nonpharmacological interventions.

With the explosive advances in technology, evidence suggests that computerized nonpharmacological interventions, including serious games, could play a positive role in improving executive functions [11]. Serious games are defined as games that are developed for specific purposes other than entertainment such as education, prevention, screening, diagnosing, and therapeutic rehabilitation [15,16]. Serious games had shown promising results in improving attention, concentration, and working memory [17]. Depending on their therapeutic modality, serious games may exist in a variety of formats, including (1) exergames, or videogames that require physical exercise as part of playing the game; (2) cognitive training games that aim to maintain or improve users' cognitive abilities (eg, executive functions, memory, learning); (3) computerized cognitive behavioral therapy (CBT) games, which are video games that provide CBT for the users; and (4) biofeedback games, which are video games that utilize electrical sensors attached to the participant to receive information about the participant's body state (eg, electrocardiogram sensors) and seek to influence some of the player's body functions (eg, heart rate). With the increasing access and ubiquity of handheld computers and smart devices, serious games continue to become more abundant via videogame consoles, personal computers, and, more recently, smartphones and tablets [11].

Research Gap and Aim

There are many studies that have examined the effectiveness of serious games in improving executive functions. Conducting systematic reviews to summarize the evidence in these studies is important to draw conclusions about the effectiveness of serious games in improving executive functions. Several systematic reviews have pooled findings of these studies. However, these reviews (1) focused on older adults without cognitive impairment [11,18-21], (2) included quasiexperiments or pilot randomized controlled trials (RCTs) [19,21-23], (3) did not assess the quality of the meta-analyzed evidence [11,19,22-24], (4) only focused on a specific type of serious game such as cognitive training games [11,20,22,24] and exergames [19,21,23], or (5) did not compare the effect of serious games with that of a specific comparator (eg, no intervention, conventional exercises, conventional cognitive activities) [11,19-24]. To address these gaps, the aim of this review was to investigate the effectiveness of serious games for improving executive functions among older adults with cognitive impairment.

XSL•FO
RenderX

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Multimedia Appendix 1) [25] were followed to conduct this systematic review and meta-analysis. The protocol for this review is registered at the International Prospective Register of Systematic Reviews (PROSPERO; ID: CRD42021272757).

Search Strategy

Search Sources

For the purpose of this review, the following 8 databases were searched: MEDLINE (via Ovid), PsycINFO (via Ovid), EMBASE (via Ovid), CINAHL (via EBSCO), IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar. Searches were completed on November 10, 2021, by the first author, and an automatic alert was set up and ran its course for 8 weeks (ending on December 5, 2021). Only the first 10 pages (ie, 100 hits) in Google Scholar were considered because it returns a large number of studies that are automatically ordered based on their relevance [26]. We conducted backward reference list checking (ie, screening of reference lists of the included studies and relevant reviews). Finally, the studies that cited the included studies were screened (ie, forward reference list checking).

Search Terms

To develop the search query, we consulted 2 experts in digital mental health. The search terms included those related to the target population (eg, cognitive disorder), target intervention (eg, serious games), and target study design (eg, RCTs). Multimedia Appendix 2 summarizes the search query that was used to search each of the 8 databases.

Study Eligibility Criteria

Only RCTs that evaluated the effectiveness of serious games for improving executive functions among older adults with cognitive impairment were included in this study. Serious games that were available on any digital platform, such as PCs, consoles (eg, Xbox, PlayStation), mobile phones, tablets, handheld devices, Nintendo, or any other computerized device, were included in this study. Furthermore, gaming had to be a key component of the intervention and used purely for therapeutic purposes. Studies combining serious games with other interventions were included if the control group received the same adjacent intervention. Nondigital games (eg, paper-and-pencil games or board games), as well as those used for monitoring, screening, diagnosis, and research, were excluded.

The population of interest was adults over 60 years old with any type of cognitive impairment or disorder (MCI, AD, or dementia). Their diagnosis had to be confirmed by checking the inclusion criteria or baseline scores against defined diagnostic criteria (eg, Mini-Mental State Examination [MMSE]). Studies about older adults without cognitive impairment, health care providers, and caregivers were excluded. No restrictions were applied regarding gender and ethnicity.

Regardless of the tool used to measure the outcome, the outcome of interest in this review was executive functions. This review did not focus on a specific executive function. Studies were excluded if they focused on only cost effectiveness, acceptance, feasibility, satisfaction, or cognitive abilities other than executive functions. This review focused on outcome data collected just after the intervention (postintervention data), rather than data collected later (follow-up data).

For practical reasons, only studies in the English language were eligible for inclusion. Although we considered all types of RCTs (parallel, cluster, crossover, and factorial) in this review, pilot RCTs, quasiexperiments, observational studies, and reviews were excluded. Research published in journals, conference proceedings, and dissertations from 2010 onwards were included. Those published as conference abstracts, conference posters, commentaries, proposals, and editorials were excluded. No restrictions related to the country of publication, comparator, and study settings were applied.

Study Selection

The following steps were followed to identify relevant studies. First, the obtained studies were imported into EndNote to identify and delete duplicate items. Second, the titles and abstracts of all retrieved studies were evaluated by 2 reviewers working independently. Third, the 2 reviewers independently checked the entire texts of the studies included in the previous step. All disagreements were resolved via discussion between the reviewers. The interrater agreements (Cohen κ) in steps 2 and 3 were 0.86 and 0.94, respectively.

Data Extraction

The 2 reviewers extracted data from included studies independently using Microsoft Excel. Before extracting data, we pilot tested the data extraction form with 2 of the included studies. Disagreements between the reviewers were settled through discussions between both reviewers. Multimedia Appendix 3 presents the data extraction form used to extract data from the included studies. First and corresponding authors were contacted in an attempt to retrieve metrics such as mean, standard deviation, and sample size if they were unavailable from the published studies.

Risk of Bias Appraisal

The 2 reviewers used the Risk-of-Bias 2 (RoB-2) tool [27] to independently appraise the risk of bias in the included studies. The RoB-2 tool evaluates the risk of bias in 5 areas of RCTs: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the result [27]. All disagreements were resolved via discussion between the reviewers. The interrater agreement between the reviewers was 0.90.

Data Synthesis

To summarize the collected data, narrative and statistical methods were used. Texts and tables were used to describe the characteristics of the included studies (demographic, intervention, comparison, and outcome measures) in our narrative synthesis. The results of the experiments were aggregated and classified by comparator: no or passive intervention control, conventional exercises, and other serious games. Meta-analyses were performed when 2 or more studies

XSL•FO
RenderX

from the same comparator submitted sufficient data (ie, mean, standard deviation, and number of participants in each intervention group). Meta-analysis was performed using Review Manager (RevMan 5.4). The standardized mean difference (SMD; Cohen *d*) was used to estimate the overall effect of each study as the type of data for the outcome of interest (executive functions) was continuous and instruments used to evaluate the outcome were diverse among the included trials. We selected a random effects model for the analysis due to the excessive clinical heterogeneity among the meta-analyzed research in terms of serious game characteristics (eg, types, duration, frequency, and period), population characteristics (eg, sample size, mean age, and health condition), and outcome measures (ie, tools and follow-up period).

If there was a statistically significant difference between the groups when performing a meta-analysis, we sought to further investigate whether it was clinically significant. The term minimal clinically important difference (MCID) refers to the smallest change in measurement results that the patient considers reasonable and important enough to justify a change in treatment. The MCID boundaries were computed as ±0.5 times the SMD of the meta-analyzed studies.

To evaluate the degree and statistical significance of heterogeneity in the meta-analyzed studies, we calculated I^2 and a chi-square *P* value, respectively. A chi-square *P* value ≤.05 suggests heterogeneous meta-analyzed studies [28]. The degree of heterogeneity was judged insignificant, moderate, substantial, or considerable when I^2 ranged from 0% to 40%, 30% to 60%, 50% to 90%, or 75% to 100%, respectively [28].

To assess the overall quality of the evidence obtained from the meta-analysis, we used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [29]. The GRADE approach appraises the quality of evidence based on 5 domains: risk of bias, inconsistency (ie, heterogeneity), indirectness, imprecision, and publication bias [29]. The 2 reviewers independently evaluated the overall quality of the meta-analyzed evidence, and the differences in decisions were addressed via discussion. The interrater agreement between the reviewers was 0.89 [30].

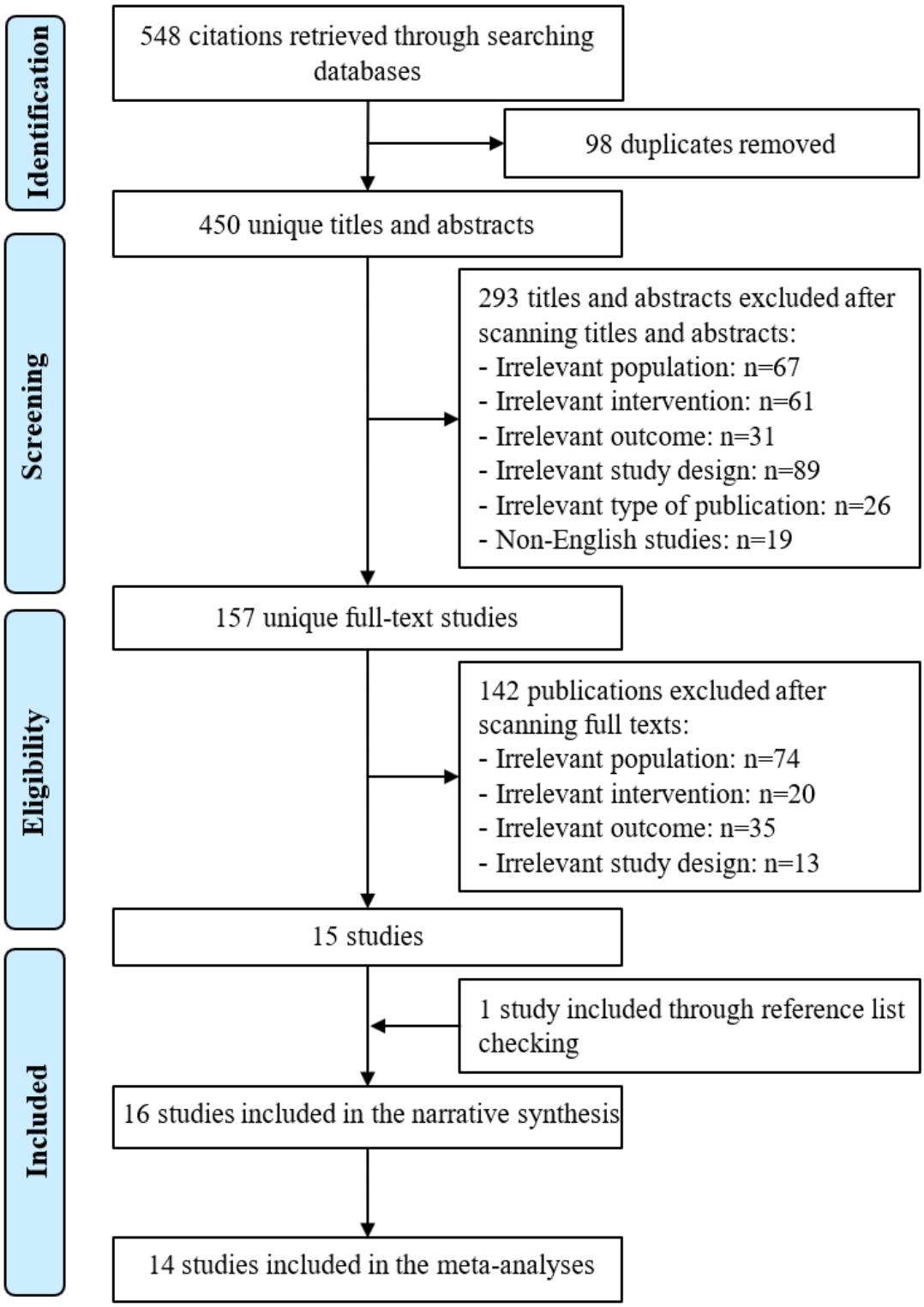
Results

Search Results

By searching the 8 electronic databases, 548 records were retrieved (Figure 1). Of these records, 98 duplicates were excluded using the software EndNote. Checking titles and abstracts of the remaining records led to excluding 293 records for the following reasons: (1) Participants were younger than 60 years and/or without cognitive impairment (n=67); (2) interventions were not serious games (n=61); (3) the outcome was not executive functions (n=31); (4) study design was not an RCT (n=89); (5) studies were not peer-reviewed articles, theses, or conference proceedings (n=26); and (6) they were published in languages other than English (n=19). Reading the full text of the remaining 157 publications led to excluding 142 publications for the following reasons: (1) Participants were younger than 60 years and/or without cognitive impairment (n=74), (2) interventions were not serious games (n=20), (3) the outcome was not executive functions (n=35), and (4) study design was not an RCT (n=13). One additional study was found through backward reference list checking. In total, 16 RCTs were included in the current review [31-46]. All studies were included in meta-analyses except 2 studies [45,46].

XSL•FO
RenderX

Figure 1. Flow chart of the study selection process.



Characteristics of Included Reviews

The included studies were published between 2013 and 2021 (Table 1). The included studies originated in 11 different countries, with a roughly equal proportion of research in each country. There was a general equal distribution of studies in these countries. Except for 1 book chapter, all included papers were peer-reviewed academic publications. The trial type used in the most included studies was parallel RCTs (n=14).

XSL•FO
RenderX

Table 1. Characteristics of studies and population.

First author	Year	Country	Publication type	RCT[a] type	Sample size	Mean age (years)	Sex (male), %	MMSE[b] score	Health condition	Setting
Cavallo [31]	2016	Italy	Journal article	Parallel	80	76.4	36.3	22.9	AD[c]	Clinical
Finn [32]	2015	Australia	Journal article	Parallel	31	75.6	71	28.1	MCI[d]	Clinical
Yang [33]	2017	South Korea	Journal article	Parallel	20	71	70	23.1	AD	Clinical
Zhuang [34]	2013	China	Journal article	Parallel	33	83.1	24.2	10.2	MCI, dementia	Clinical
Thapa [35]	2020	South Korea	Journal article	Parallel	68	72.7	23.5	26.2	MCI	Clinical
Tarnanas [36]	2014	Greece	Book chapter	Parallel	114	70.3	39	26.4	MCI	Clinical
Singh [37]	2014	Australia	Journal article	Factorial	100	70.1	32	27	MCI	Community
Amjad [38]	2019	Pakistan	Journal article	Parallel	44	NR[e]	NR	24	MCI	Clinical
Hagovská [39]	2016	Slovakia	Journal article	Parallel	80	67	51.2	26.4	MCI	Clinical
van Santen [40]	2020	Netherlands	Journal article	Cluster	112	79	53.5	18.6	Dementia	Clinical
Karssemeijer [41]	2019	Netherlands	Journal article	Parallel	115	79.9	53.9	22.4	Dementia	Clinical & community
Liao [42]	2021	Taiwan	Journal article	Parallel	61	81.5	32.6	22.9	MCI	Community
Flak [43]	2019	Norway	Journal article	Parallel	85	66	66.7	NR	MCI	Clinical
Hyer [44]	2016	United States	Journal article	Parallel	68	75.2	47.1	26	MCI	Community
Park [45]	2017	South Korea	Journal article	Parallel	78	67.3	53.8	26.5	MCI	Community
Lee [46]	2018	South Korea	Journal article	Parallel	20	74.3	40	17.9	MCI, AD, dementia	Clinical

[a]RCT: randomized controlled trial.

[b]MMSE: Mini-Mental State Examination.

[c]AD: Alzheimer disease.

[d]MCI: mild cognitive impairment.

[e]NR: not reported.

The sample size in the included studies ranged from 20 to 115, with an average of 69.3. The average age of participants in the 15 studies was 74 years, with a range of 67 years to 83 years. The percentage of men reported in 15 studies ranged from 24.2% to 71%, with an average of 46.3%. Participants in the included studies had a mean MMSE score of 23.2, with a range of 10.2 to 28.1. Participants in the included studies had MCI (n=10), AD (n=2), dementia (n=2), MCI and dementia (n=1), and all 3 (MCI, AD, and dementia; n=1). Participants were recruited from clinical settings (n=11), the community (n=4), and clinical and community settings (n=1).

In 14 of the studies considered, serious games were employed alone as therapies, whereas the other 2 studies combined serious games with other interventions (Table 2). We identified 18 distinct serious games used in the studies; more than 1 game was used in certain studies. Serious games in the included trials were divided into 2 categories depending on the treatment modality they provide: cognitive training games (n=13) and exergames (n=3). In 14 studies, games were created with a "serious" objective from the start (designed serious games). Games in the remaining 2 studies, on the other hand, were not planned as serious games from the start but were instead used for a serious purpose (purpose-shifted games). Computers were the most popular platforms for playing games in the included studies (n=10). In most studies (n=11), serious games were played under the supervision of health care providers or carers. The game durations in the included studies ranged from 25 minutes to 100 minutes. The frequency of playing the games ranged from 2 times to 7 times per week, but it was 3 times per week in roughly one-third (6/16, 38%) of the studies. The duration of interventions ranged from 3 weeks to 25 weeks but was less than 13 weeks in the majority of studies (12/16, 75%).

The comparison groups in 7 studies received no or passive interventions (eg, reading newspaper articles, surfing the internet, watching a documentary program), whereas active interventions (eg, conventional exercises, other serious games) were conducted in 8 studies (Table 3). Two studies delivered both active as well as passive interventions as comparators. The duration of the active comparators ranged between 25 minutes and 100 minutes. The active comparators were used between once a week and 7 times a week. The duration of the active comparators ranged from 4 weeks to 25 weeks. The outcome

XSL•FO
RenderX

of interest (ie, executive functions) was measured using 13 different tools, but the Trail Making Test B (TMT-B) was the most commonly used tool by the included studies (8/16, 50%). In all included studies, the outcome of interest was measured immediately after the intervention, and the longest follow-up period was 74 weeks. The number of participants who dropped out ranged from 0 to 28.

Table 2. Characteristics of interventions.

First author	Intervention	Serious game name	Serious game type	Serious game genre	Platform	Supervision	Duration (minutes)	Frequency (times/week)	Period (weeks)
Cavallo [31]	Serious games	Brainer	Cognitive training game	Designed	PC	Supervised	30	3	12
Finn [32]	Serious games	E-Prime	Cognitive training game	Designed	PC	Supervised	NR[a]	2	4
Yang [33]	Serious games	Brain-Care	Cognitive training game	Designed	PC	Unsupervised	60	2	12
Zhuang [34]	Serious games + sham exercises	NR	Cognitive training game	Designed	PC	Supervised	75	3	24
Thapa [35]	Serious games	Juice making, Crow Shooting, Love house, Fireworks	Cognitive training game	Designed	VR[b] headset, hand controllers	Supervised	100	3	8
Tarnanas [36]	Serious games	Virtual Reality Museum	Cognitive training game	Designed	VR headset	Supervised	90	2	21
Singh [37]	Serious games	COGPACK	Cognitive training game	Designed	PC	Supervised	75	2	25
Amjad [38]	Serious games	Body and Brain Exercises	Cognitive training game	Purpose-shifted	Xbox console, Kinect	Supervised	25-30	5	6
Hagovská [39]	Serious games + conventional exercises	CogniPlus	Cognitive training game	Designed	PC	Supervised & unsupervised	30	2	10
van Santen [40]	Serious games	NR	Exergame	Designed	Stationary bike & screen	Unsupervised	NR	5	25
Karssemeijer [41]	Serious games	NR	Exergame	Purpose-shifted	Stationary bike & screen	Supervised	30-50	3	12
Liao [42]	Serious games	Tano and Long-Good	Exergame	Designed	Kinect, VR headset	Supervised	60	3	12
Flak [43]	Serious games	Cogmed	Cognitive training game	Designed	PC	Unsupervised	30-40	5	5
Hyer [44]	Serious games	Cogmed	Cognitive training game	Designed	PC	Supervised & unsupervised	40	7	5-7
Park [45]	Serious games	CoTras	Cognitive training game	Designed	PC	Supervised	30	3	10
Lee [46]	Serious games	Bettercog	Cognitive training game	Designed	PC	Supervised	30	4	3

[a]NR: not reported.

[b]VR: virtual reality.

XSL•FO
RenderX

Table 3. Characteristics of comparators and outcomes.

First author	Comparator	Duration (minutes)	Frequency (times/week)	Period (weeks)	Outcome measures	Follow-up	Attrition, n
Cavallo [31]	Control	NA[a]	NA	NA	HSCT[b], LVF[c], BT[d]	Postintervention, 24-week follow-up	4
Finn [32]	Control	NA	2	4	D-KEFS[e]	Postintervention	7
Yang [33]	Control	NA	NA	NA	COWAT[f]	Postintervention	0
Zhuang [34]	Control	NA	NA	NA	ACE-R-F[g]	Postintervention	10
Thapa [35]	Control	30-50	1	8	TMT-B[h]	Postintervention	2
Tarnanas [36]	1: control; 2: conventional cognitive activities	90	2	21	TMT-B, LVF	Postintervention	9
Singh [37]	1: conventional exercises + sham cognitive training; 2: serious games + conventional exercises; 3: control	1: 75; 2: 100; 3: 60	2	25	COWAT, CF[i], WAIS-III-S[j], WAIS-III-M[k]	Postintervention, 74-week follow up	14
Amjad [38]	Conventional exercises	25-30	5	6	TMT-B	Postintervention	6
Hagovská [39]	Conventional exercises	30	7	10	ACE-WP[l]	Postintervention	2
van Santen [40]	Conventional exercises	NA	5	25	TMT-B	Midintervention, postintervention	28
Karssemeijer [41]	1: conventional exercises (aerobic exercises); 2: conventional exercises (relaxation and flexibility exercises)	30-50	3	12	TMT-B, LVF, RSCT[m]	Midintervention, postintervention, 24-week follow-up	23
Liao [42]	Conventional exercises	60	3	12	TMT-B, EXIT-25[n]	Postintervention	15
Flak [43]	Nonadaptive serious game	30-40	5	5	D-KEFS-CWIT3[o], D-KEFS-CWIT4[p], D-KEFS-VFTLF[q], D-KEFS-VFTCF[r], D-KEFS-VFTCS[s]	Postintervention, 4-week follow-up, 16-week follow-up	17
Hyer [44]	Nonadaptive serious game	40	7	5-7	TMT-B	Postintervention, 12-week follow-up	9
Park [45]	Serious game (exergames)	30	3	10	TMT-B	Postintervention	0
Lee [46]	Serious game (targeting attention and memory)	30	4	3	SNSB-II[t]	Postintervention	1

[a]NA: not applicable.

[b]HSCT: Hayling Sentence Completion Test.

[c]LVF: latter verbal fluency.

[d]BT: Brixton test.

[e]D-KEFS: Delis–Kaplan Executive Function System.

[f]COWAT: Controlled Oral Word Association Test.

[g]ACE-R-F: Addenbrooke's Cognitive Examination-Revised-fluency.

[h]TMT-B: Trail Making Test B.

[i]CF: category fluency.

[j]WAIS-III-S: Wechsler Adult Intelligence Scale III-Similarities.

[k]WAIS-III-M: Wechsler Adult Intelligence Scale III-Matrices.

[l]ACE-WP: Addenbrooke's Cognitive Examination-Word production.

XSL•FO
RenderX

[m]RSCT: Rule Shift Cards Test.

[n]EXIT-25: Executive Interview 25.

[o]D-KEFS-CWIT3: Delis–Kaplan Executive Function System-Color Word Interference Test 3.

[p]D-KEFS-CWIT4: Delis–Kaplan Executive Function System-Color Word Interference Test 4.

[q]D-KEFS-VFTLF: Delis–Kaplan Executive Function System-Verbal Fluency Test Letter Fluency.

[r]D-KEFS-VFTCF: Delis–Kaplan Executive Function System-Verbal Fluency Test Category Fluency.

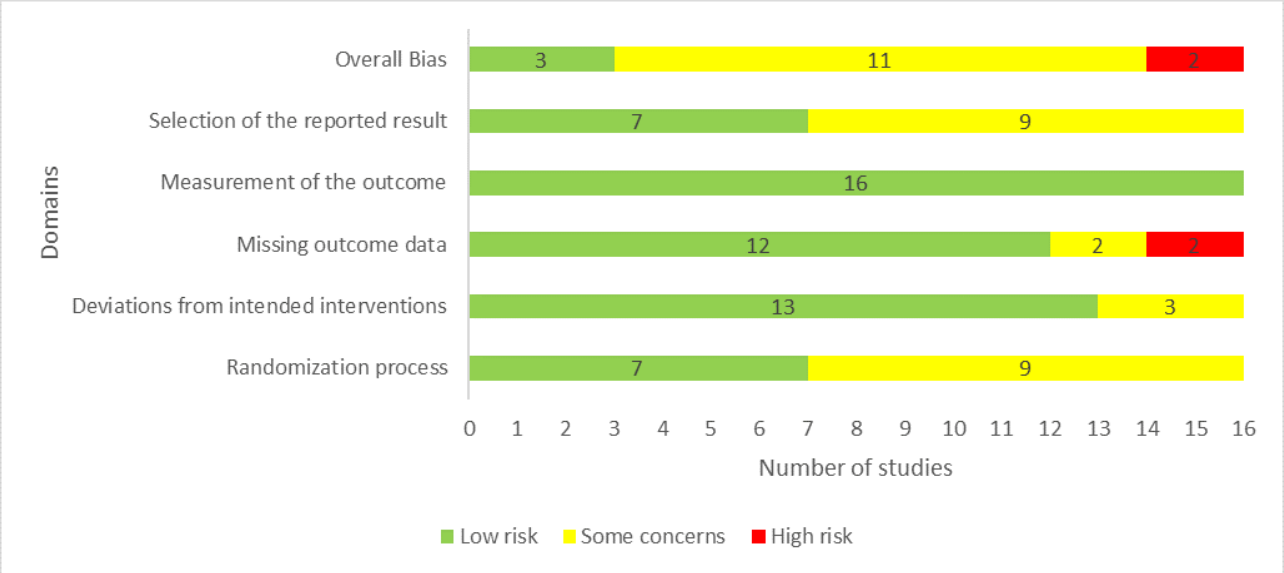
[s]D-KEFS-VFTCS: Delis–Kaplan Executive Function System-Verbal Fluency Test Category Switching.

[t]SNSB-II: Seoul Neuropsychological Screening Battery 2nd edition.

Results of Risk of Bias Appraisal

In 10 studies, participants were randomly assigned to groups using a proper random allocation sequence. Half of the included studies concealed the allocation sequence until participants were assigned to interventions. Groups were comparable at baseline in 15 studies. Accordingly, 7 of the 16 studies were judged to have a low risk of bias in the "randomization process" domain (Figure 2).

Figure 2. Review authors' judgments about each "Risk of bias" domain.



Participants were aware of the assigned interventions during the trial in 12 studies. In 14 studies, individuals who delivered the interventions to the participants were aware of the assigned interventions. There was no evidence that the experimental contexts led to a deviation from the intended intervention in 14 studies. All included studies except for 1 study estimated the effect of the intervention using appropriate analysis methods (eg, intention-to-treat analysis). Consequently, 13 of the 16 studies were judged to have a low risk of bias in the "deviations from the intended interventions" domain (Figure 2).

In 6 studies, outcome data were available for more than 95% of the participants. In only 1 study, there was evidence that the findings were not biased by missing outcome data. The missing outcome data could be related to participants' health status in 3 studies. According to these judgments, the risk of bias due to missing outcome data was low in 12 studies (Figure 2).

In all included studies, executive function was examined using appropriate measures, and measurement methods were comparable across intervention groups. In 6 studies, the assessor of the outcome was aware of the assigned interventions. In all studies, assessment of the outcome may not have been affected by knowledge of the intervention received. Therefore, the risk of bias in the "measuring the outcome" domain was rated as low in all studies (Figure 2).

Of the studies, 7 published their protocol in sufficient detail. In all studies, reported outcome measurements did not differ from those specified in the analysis plan, and there is no evidence that studies selected their results from many results produced from multiple eligible analyses of the data. Based on these judgments, 7 studies were judged to have a low risk of bias in the "selection of the reported results" domain (Figure 2).

In the last domain, "overall bias," 3 studies were judged to be at low risk of bias given that it was rated to be at low risk of bias for all other domains. Because they had some issues in at least one of the domains and were not at high risk for any domain, 11 studies raised some concerns in the domain of overall bias. The risk of bias was rated high in 2 studies, as they were judged as having a high risk of bias in at least one domain. Reviewers' judgments about each "risk of bias" domain for each included study are presented in Multimedia Appendix 4.

Results of Studies

Serious Games Versus No or Passive Interventions

In 7 studies [31-37], the effect of serious games was compared with a control (no or passive intervention). Passive interventions refer to interventions that do not have a known effect on the measured outcome such as reading newspaper articles, surfing the internet, and watching a documentary program. Of these

XSL•FO
RenderX

studies, 4 assessed executive function using more than one measure [31-33,36]. Therefore, we included the results of all these measures in the meta-analysis to form 15 comparisons (Figure 3). The meta-analysis showed no statistically significant difference (P=.29) in executive function between serious games and control groups (SMD –0.19, 95% CI –0.54 to 0.16). The statistical heterogeneity of the evidence was considerable (P<.001, I^2=81%). The quality of the evidence was very low, as it was downgraded by 6 levels due to high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5).

The SMD of 2 comparisons seemed to be outliers (–2.15 [36] and 0.81 [31]), although characteristics of the studies in these comparisons were comparable to the other studies in this meta-analysis. Thus, we conducted a sensitivity analysis to check whether removing these outliers influenced the overall effect size and heterogeneity level. The sensitivity analysis showed that the difference in executive function between the groups remained insignificant (P=.17), but the heterogeneity substantially decreased from 81% to 31%.

We conducted a subgroup analysis to assess whether the effect of serious games is based on the health conditions of participants. As shown in Figure 4, there was a statistically significant difference (P=.002) between the effect of serious games on executive functions among older adults with MCI (SMD 0.33) and their effect on executive functions among older adults with dementia (SMD 0.20) when compared with a control.

Figure 3. Forest plot of 7 studies (15 comparisons) comparing the effect of serious games to control on executive functions.

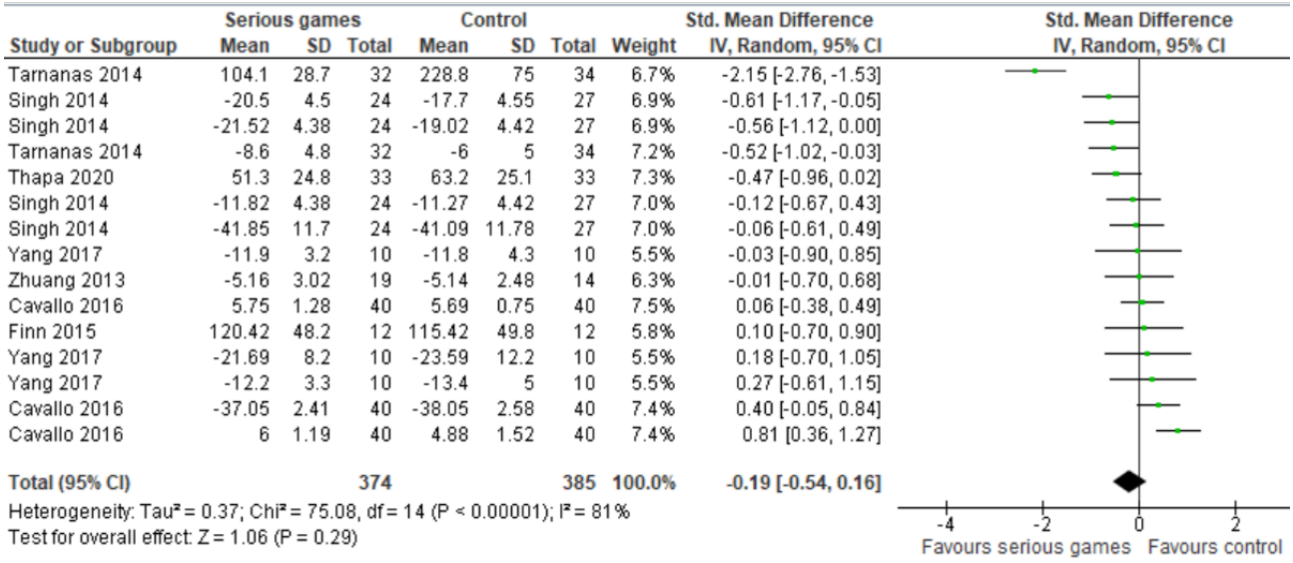
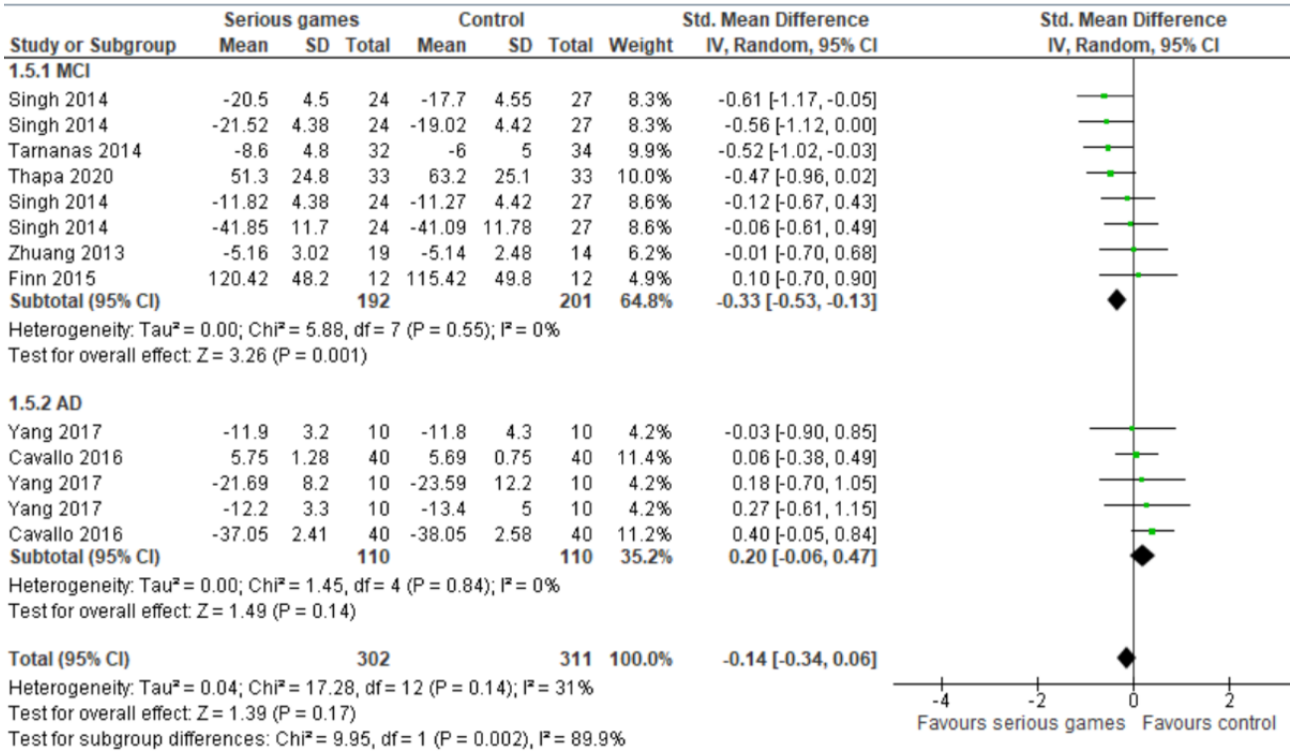


Figure 4. Forest plot of 7 studies (13 comparisons) comparing the effect of serious games on older adults with mild cognitive impairment (MCI) to their effect on older adults with Alzheimer disease (AD).

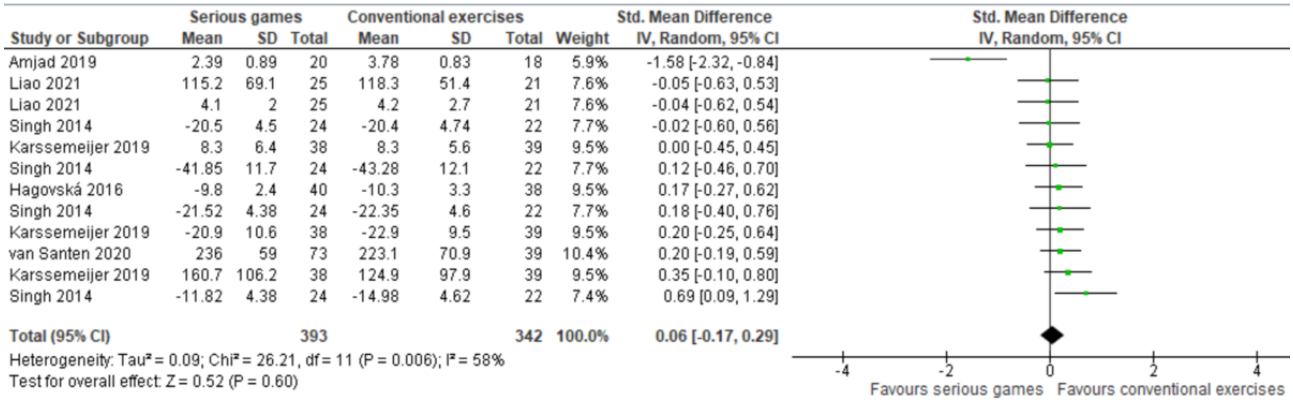


XSL•FO
RenderX

Serious Games Versus Conventional Exercises

In 6 studies [37-42], the effect of serious games was compared with conventional exercises. Of these studies, 3 evaluated executive function using more than one measure [37,41,42]. Therefore, we included the results of all these measures in the meta-analysis to form 12 comparisons (Figure 5). The meta-analysis showed no statistically significant difference (*P*=.60) in executive function between the serious games group and conventional exercises group (SMD 0.06, 95% CI –0.17 to 0.29). The statistical heterogeneity of the evidence was moderate (*P*=.006, I^2=58%). The quality of the evidence was very low, as it was downgraded by 5 levels due to high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5).

Figure 5. Forest plot of 6 studies (12 comparisons) comparing the effect of serious games to conventional exercises on executive functions.



The effect size in 1 study [38] seemed to be an outlier (–1.58). This could be attributed to the following reasons: (1) The sample size in this study was the smallest (n=44) in all the meta-analyzed studies, (2) it was the only study that used a videogame console (Xbox) as a platform for the serious game, and (3) the interventions in this study were delivered for a short period (6 weeks) in comparison with other studies included in this meta-analysis. Accordingly, we ran a sensitivity analysis to check whether removing this outlier influenced the overall effect size and heterogeneity level. The sensitivity analysis showed a statistically significant difference in executive functions (*P*=.03) between the groups, favoring conventional exercises over serious games (SMD 0.17, 95% CI 0.02 to 0.32). This difference was also clinically important, as the overall effect was outside the MCID boundaries (–0.085 to 0.085) and its CI did not cross the "no effect" line (zero effect). For this outcome, MCID boundaries were calculated as ±0.5 times the SMD value (0.17). The statistical heterogeneity of the evidence was not a concern (*P*=.85, I^2=0%). The quality of this evidence was very low, as it was downgraded by 3 levels due to high risk of bias and imprecision.

In this comparison (ie, serious games vs conventional exercises), 2 types of serious games were used: cognitive training games and exergames. We conducted a subgroup analysis to investigate whether different types of serious games (ie, cognitive training games and exergames) have a different effect on executive functions (Figure 6). The subgroup analysis showed no statistically significant difference (*P*=.61) between cognitive training games (SMD 0.22) and exergames (SMD 0.14) in their effect on executive functions when compared with conventional exercises.

Further, we conducted a subgroup analysis to assess whether the effect of serious games is based on the health conditions of participants. As shown in Figure 7, there was no statistically significant difference (*P*=.80) between the effect of serious games on executive functions among older adults with MCI (SMD 0.15) and their effect on executive functions among older adults with dementia (SMD 0.19) when compared with conventional exercises.

XSL•FO
RenderX

Figure 6. Forest plot of 5 studies (11 comparisons) comparing the effect of cognitive training games and exergames to conventional exercises.

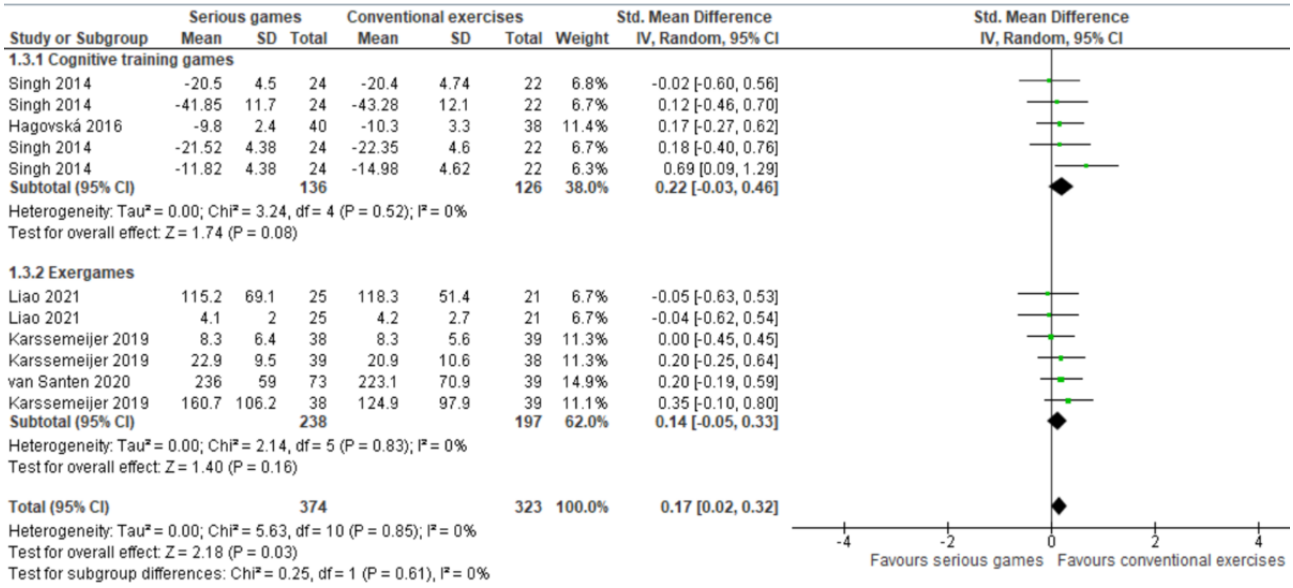
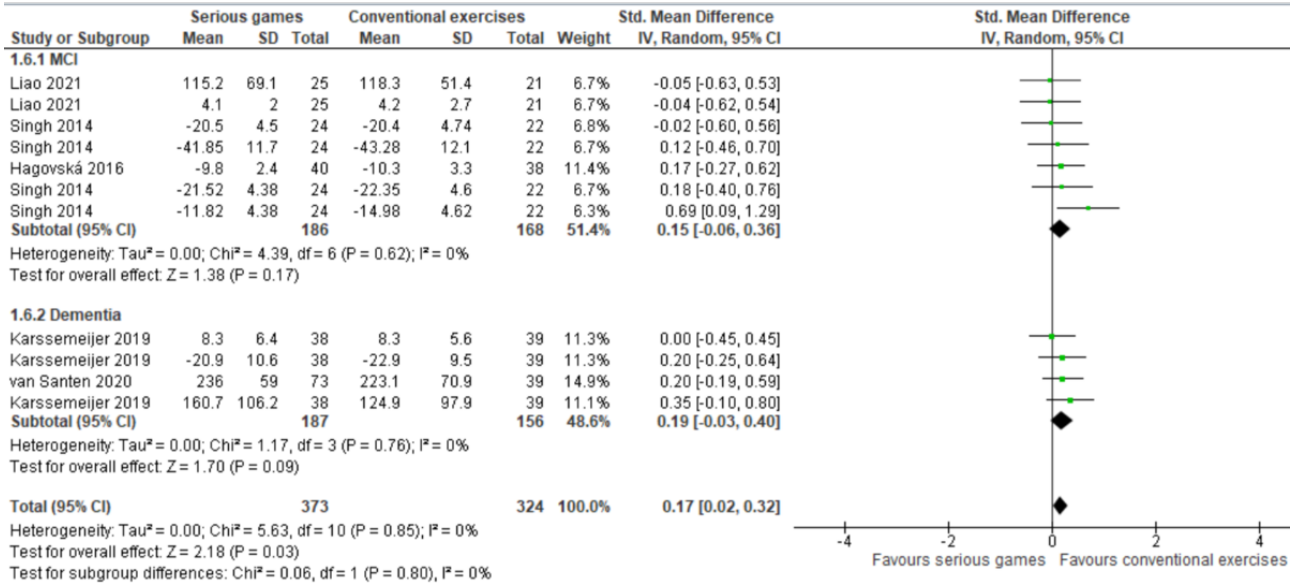


Figure 7. Forest plot of 5 studies (11 comparisons) comparing the effect of serious games on older adults with mild cognitive impairment (MCI) to their effect on older adults with dementia.



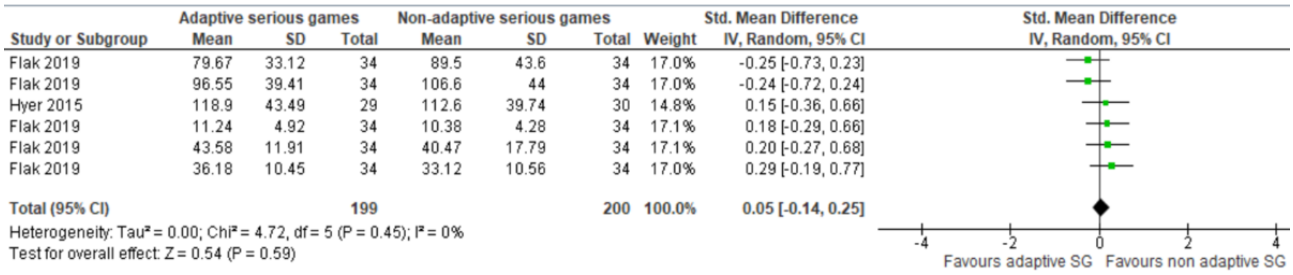
Serious Games Versus Other Serious Games

The effect of serious games on executive function in comparison with other serious games was assessed in 4 studies [43-46]. Specifically, the first study compared the effect of a cognitive training game with exergames [45]. The study showed no statistically significant difference (*P*=.52) in executive functions between the groups [45]. The second study compared the effect of a cognitive training game that targets only memory and attention (COMCOG) with another cognitive training game that targets many cognitive abilities (ie, orientation, attention, memory, language, executive function, visuospatial function, calculation, and motor functions; Bettercog) [46]. This study found no statistically significant difference (*P*=.07) in executive functions between the groups [46].

The 2 remaining studies compared the effect of cognitive training games that adjust the level of difficulty of the tasks based on the individual's mastery in each level (ie, adaptive games) with the same games but without adjustment of the level of difficulty of the tasks (ie, nonadaptive game) [43,44]. One of these studies assessed executive function using 4 different measures [43]. Thus, we ran a meta-analysis using these measures to form 5 comparisons. As shown in Figure 8, there was no statistically significant difference (*P*=.59) in executive functions between groups (SMD 0.05, 95% CI –0.14 to 0.25). The statistical heterogeneity of the evidence was not a concern (*P*=.45, I^2=0%). The quality of the evidence was very low, as it was downgraded by 3 levels due to high risk of bias and imprecision (Multimedia Appendix 5).

XSL•FO
RenderX

Figure 8. Forest plot of 2 studies (5 comparisons) comparing the effect of adaptive serious games to nonadaptive serious games.



Discussion

Principal Findings

This study summarized the evidence about the effectiveness of serious games for improving executive functions among older adults with cognitive impairment. Our meta-analyses showed that serious games are as effective as no or passive interventions at improving executive functions. Surprisingly, we found that conventional exercises are more effective than serious games at improving executive functions. However, our subgroup analysis showed that cognitive training games and exergames have a comparable effect on executive functions and they are as effective as conventional exercises at improving executive functions. We also found no difference between adaptive serious games and nonadaptive serious games at improving executive functions.

The lack of superiority of serious games over no or passive interventions and conventional exercises may be attributed to the following reasons: (1) The content of serious games in the meta-analyzed studies did not specifically target executive functions; (2) the target population (older adults with cognitive impairment) are less likely to be able to effectively play serious games; (3) most included studies assessed overall executive function rather than specific executive functions (eg, inhibition, impulse control, self-monitoring, task initiation, emotional control, flexible thinking), leaving it unclear whether the effect of serious games is different for different executive functions; and (4) the sample size in most included studies was small (≤100).

The findings of our review and those of previous reviews were consistent for some comparisons and different for others. Specifically, Lampit et al [20] summarized the evidence about the effect of cognitive training games on executive functions among healthy older adults in comparison with passive and active interventions. The review found cognitive training games are as effective as active interventions in improving executive functions; however, it showed that serious games are more effective at improving executive functions in comparison with no or passive interventions [20]. Another review compared the effect of cognitive training games on executive functions with any other interventions (passive or active interventions) [11]. The review demonstrated that cognitive training games are more effective than other interventions at improving executive functions among healthy older adults [11]. Our findings are inconsistent with the findings of these reviews [11,20]. This difference may be attributed to the fact that both reviews focused on healthy older adults only while the current review focused on older adults with cognitive impairment.

Meta-analyses from 2 other reviews showed that the effect of cognitive training games on executive functions among older adults with cognitive impairment is not statistically significant in comparison with other passive and active interventions [22,24]. Our findings are in line with the results of these reviews [22,24]. However, the main differences between the current review and these reviews are as follows: (1) The previous reviews focused only on a specific type of serious games (ie, cognitive training games), while the current review focused on all types of serious games; (2) they did not compare the effect of cognitive training games with a specific type of comparator (no or passive interventions, conventional exercise, other serious games); and (3) they included pilot RCTs and quasiexperiments, whereas the current review excluded such studies.

A systematic review conducted by Yen and Chiu [21] showed that exergames do not significantly improve executive functions among older adults in comparison with other passive and active interventions. Our findings are consistent with the finding of the previous review [21]. In contrast, another review found that exergames are more effective than passive and active interventions in improving executive functions among healthy older adults [19]. This contradictory finding may be attributed to 2 reasons: (1) Although the former review [21] focused on older adults with and without cognitive impairment, the latter review [19] focused on the older adults without cognitive impairment, and (2) the former review [21] assessed the effect of virtual reality exergames, while the latter review [19] examined the effect of exergames in general.

None of the previous reviews compared the effect of adaptive serious games with nonadaptive serious games on executive functions. However, a review compared the effect of adaptive serious games with that of nonadaptive serious games on working memory among older adults with cognitive impairment [47]. The review found no statistically significant difference in working memory between groups [47], and this is in line with our findings.

Strengths and Limitations

Strengths

In comparison with previous reviews [11,18-24], this review is the first of its kind, to the best of our knowledge, that compares both the effect of serious games and their types on executive functions with a specific comparator (ie, no intervention, conventional exercises, and other serious games). Further, this review is the first of its kind to use the GRADE approach to

XSL•FO
RenderX

appraise the quality of the evidence resulting from the meta-analyses, and this enables the reader to draw more accurate conclusions.

This review followed highly recommended guidelines for reporting systematic reviews (ie, PRISMA); thus, it can be considered a transparent and reproducible review. Our findings are based on RCTs, which are the most rigorous research method in studying cause-effect relationships [48]. Hence, the findings of this review are more likely to be reliable than findings generated from reviews that included other study designs such as pilot RCTs and quasiexperiments.

The risk of publication bias in this current review is not a concern, as the authors sought to retrieve as many relevant studies as possible through searching the most popular databases in information technology and health fields and grey literature databases, conducting backward and forward reference list checking, and using a well-developed search query. In addition, the risk of selection bias in this review is minimal because the study selection, data extraction, risk of bias assessment, and quality of evidence appraisal were conducted by 2 reviewers independently.

Limitations

The current review focused on the effectiveness of digital serious games in improving executive functions among older adults with cognitive impairment. For this reason, this review cannot comment on the effectiveness (1) of nondigital serious games or those used for nontherapeutic purposes (eg, screening or diagnosis), (2) at improving a specific executive function or other cognitive abilities (eg, attention, processing speed, memory), and (3) among other age groups or those without cognitive impairment.

In this review, the effect size for each study was estimated using postintervention data rather than the pre-post intervention change for each group; thereby, it is likely that the effect size is overestimated or underestimated. Postintervention outcome data were used because most studies did not report the mean and standard deviation for pre-post intervention change in executive functions for each group and the difference in executive functions between groups at baseline was not statistically significant in all studies.

This review assessed only the short-term effect of serious games by pooling only postintervention data rather than follow-up data, as the follow-up period was not consistent between the 5 studies that reported follow-up data. Thus, we cannot comment on the long-term effect of serious games on executive functions. It is likely that this review missed some relevant studies given that we excluded studies that were published before 2010, written in a language other than English, quasiexperiments, and pilot RCTs.

Practical and Research Implications

Practical Implications

This review showed no superior effect of serious games compared with no or passive interventions and conventional exercises on executive functions among older adults with cognitive impairment. Further, there was no difference between adaptive serious games and nonadaptive serious games at improving executive functions among older adults with cognitive impairment. However, readers should cautiously interpret our findings for the following reasons: (1) The quality of evidence ranged between very low to low due mainly to high risk of bias, high heterogeneity, and imprecision of the estimated total effect sizes; (2) the number of studies included in several meta-analyses was small; and (3) the sample size in most studies included in the meta-analyses was small. Accordingly, serious games should not be offered or used for improving executive functions among older adults with cognitive functions until more robust evidence is available. This is a call to action for researchers, clinicians, and game developers to continue improving their work and focus on addressing the limitations and concerns discussed earlier.

Smart mobile devices (ie, tablets and smartphones) were not used in any study included in this review. Smart mobile devices are particularly appealing, as they are cheaper, more accessible, and more pervasive than computers and gaming consoles. Globally, the number of mobile devices and mobile users in 2021 were about 15 billion and 7.1 billion, respectively, and these figures are expected to rise considerably by 2025 [49]. There is an opportunity for smart device app developers as well as serious game developers to create and tailor serious games that target executive functions of older adults with cognitive impairment and can be played via mobile devices.

Research Implications

This review addressed the research gap related to the short-term effect of serious games on executive functions among older adults with cognitive impairment. However, further reviews are needed to address the following research gaps: (1) the long-term effect of serious games, (2) the effect of serious games on specific executive functions (eg, inhibition, impulse control, self-monitoring, flexible thinking) and on other cognitive abilities (eg, attention, processing speed, learning), and (3) the effect of serious games among people of different age groups with or without cognitive impairment.

Most included studies were conducted in developed countries; thereby, the generalizability of this review's findings to developing countries may be limited given the varying nature of their cultures and socioeconomic conditions. Researchers should carry out more studies in developing countries. The mean and standard deviation for pre-post intervention change in executive functions for each group were not reported by most of the included studies. To calculate a more accurate effect size for each study, we urge researchers to report such information in their future publications.

Previous reviews showed that the effect of exergames on executive functions among healthy older adults was investigated by many studies [19,21,23]. However, in the current review, only 3 studies examined the effect of exergames on executive functions among older adults with cognitive impairment. Further studies are required to bridge this research gap. In this review, serious games were compared with conventional cognitive training by only 1 study, and adaptive serious games were compared with nonadaptive serious games by only 2 studies.

XSL•FO
RenderX

To draw more definitive conclusions, these comparisons should be examined by further trials.

Only 3 of the included studies were judged to have a low overall risk of bias, as the remaining studies had issues mainly in the randomization process or selection of the reported results (ie, unpublished protocol or analysis plan). To minimize the risk of bias, researchers should conduct and report their trials according to recommended guidelines or tools such as the RoB-2 [27].

Conclusion

The evidence from this review showed no superior effect of serious games compared with no or passive interventions and conventional exercises on executive functions among older adults with cognitive impairment. However, this should not be considered a definitive conclusion for the following reasons: (1) The quality of evidence ranged between very low to low due mainly to high risk of bias, high heterogeneity, and imprecision of the estimated total effect sizes; (2) the number of studies included in several meta-analyses was small; and (3) the sample size in most studies included in the meta-analyses was small. Therefore, until more robust evidence is available, serious games should not be offered by health care providers nor used by patients for improving executive functions among older adults with cognitive impairment. Further reviews are needed to assess the long-term effect of serious games on specific executive functions or other cognitive abilities among people of different age groups with or without cognitive impairment. Additional RCTs should be conducted to examine the effect of exergames on executive functions among older adults with cognitive impairment and to compare the effect of serious games with conventional exercises.

Conflicts of Interest

None declared

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 26 KB - games_v10i3e36123_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 30 KB - games_v10i3e36123_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 25 KB - games_v10i3e36123_app3.docx]

Multimedia Appendix 4
Reviewers' judgments about each "risk of bias" domain for each included study.
[DOCX File , 54 KB - games_v10i3e36123_app4.docx]

Multimedia Appendix 5
GRADE Profile for comparison of serious games to control, conventional exercises, and non-adaptive serious games for executive functions.
[DOC File , 18 KB - games_v10i3e36123_app5.doc]

References

1. Grande G, Qiu C, Fratiglioni L. Prevention of dementia in an ageing world: Evidence and biological rationale. Ageing Res Rev 2020 Dec;64:101045 [FREE Full text] [doi: 10.1016/j.arr.2020.101045] [Medline: 32171784]
2. Beard JR, Araujo de Carvalho I, Sumi Y, Officer A, Thiyagarajan JA. Healthy ageing: moving forward. Bull World Health Organ 2017 Nov 01;95(11):730-730A [FREE Full text] [doi: 10.2471/BLT.17.203745] [Medline: 29147049]
3. Hebert LE, Scherr PA, Beckett LA, Albert MS, Pilgrim DM, Chown MJ, et al. Age-specific incidence of Alzheimer's disease in a community population. JAMA 1995 May 03;273(17):1354-1359. [Medline: 7715060]
4. Calderón-Larrañaga A, Vetrano DL, Onder G, Gimeno-Feliu LA, Coscollar-Santaliestra C, Carfí A, et al. Assessing and measuring chronic multimorbidity in the older population: a proposal for its operationalization. J Gerontol A Biol Sci Med Sci 2017 Oct 01;72(10):1417-1423 [FREE Full text] [doi: 10.1093/gerona/glw233] [Medline: 28003375]
5. Thakur R, Banerjee A, Nikumb V. Health problems among the elderly: a cross-sectional study. Ann Med Health Sci Res 2013 Jan;3(1):19-25 [FREE Full text] [doi: 10.4103/2141-9248.109466] [Medline: 23634324]
6. Thoma-Lürken T, Bleijlevens M, Lexis M, de Witte LP, Hamers J. Facilitating aging in place: A qualitative study of practical problems preventing people with dementia from living at home. Geriatr Nurs 2018;39(1):29-38 [FREE Full text] [doi: 10.1016/j.gerinurse.2017.05.003] [Medline: 28624128]

XSL•FO
RenderX

7. Akila GV, Arvind BA, Isaac A. Comparative assessment of psychosocial status of elderly in urban and rural areas, Karnataka, India. J Family Med Prim Care 2019 Sep;8(9):2870-2876 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_536_19] [Medline: 31681658]
8. Chen Y, Liang N, Li X, Yang S, Wang Y, Shi N. Diagnosis and treatment for mild cognitive impairment: a systematic review of clinical practice guidelines and consensus statements. Front Neurol 2021;12:719849 [FREE Full text] [doi: 10.3389/fneur.2021.719849] [Medline: 34712197]
9. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. Alzheimers Dement 2020 Mar 10:391-460 [FREE Full text] [doi: 10.1002/alz.12068] [Medline: 32157811]
10. Cimler R, Maresova P, Kuhnova J, Kuca K. Predictions of Alzheimer's disease treatment and care costs in European countries. PLoS One 2019 Jan 25;14(1):e0210958 [FREE Full text] [doi: 10.1371/journal.pone.0210958] [Medline: 30682120]
11. Bonnechère B, Langley C, Sahakian BJ. The use of commercial computerised cognitive games in older adults: a meta-analysis. Sci Rep 2020 Sep 17;10(1):15276 [FREE Full text] [doi: 10.1038/s41598-020-72281-3] [Medline: 32943742]
12. Diamond A. Executive functions. Annu Rev Psychol 2013;64:135-168 [FREE Full text] [doi: 10.1146/annurev-psych-113011-143750] [Medline: 23020641]
13. Benzing V, Schmidt M. The effect of exergaming on executive functions in children with ADHD: A randomized clinical trial. Scand J Med Sci Sports 2019 Aug;29(8):1243-1253 [FREE Full text] [doi: 10.1111/sms.13446] [Medline: 31050851]
14. Cristofori I, Cohen-Zimerman S, Grafman J. Executive functions. Handb Clin Neurol 2019;163:197-219. [doi: 10.1016/B978-0-12-804281-6.00011-2] [Medline: 31590731]
15. Manera V, Ben-Sadoun G, Aalbers T, Agopyan H, Askenazy F, Benoit M, et al. Recommendations for the use of serious games in neurodegenerative disorders: 2016 Delphi panel. Front Psychol 2017 Jul 25;8:1243 [FREE Full text] [doi: 10.3389/fpsyg.2017.01243] [Medline: 28790945]
16. Wiemeyer J, Kliem A. Serious games in prevention and rehabilitation—a new panacea for elderly people? Eur Rev Aging Phys Act 2011 Dec 08;9(1):41-50. [doi: 10.1007/s11556-011-0093-x]
17. Chicchi Giglioli IA, de Juan Ripoll C, Parra E, Alcañiz Raya M. EXPANSE: A novel narrative serious game for the behavioral assessment of cognitive abilities. PLoS One 2018;13(11):e0206925 [FREE Full text] [doi: 10.1371/journal.pone.0206925] [Medline: 30412614]
18. Wang G, Zhao M, Yang F, Cheng LJ, Lau Y. Game-based brain training for improving cognitive function in community-dwelling older adults: A systematic review and meta-regression. Arch Gerontol Geriatr 2021;92:104260. [doi: 10.1016/j.archger.2020.104260] [Medline: 32980574]
19. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
20. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
21. Yen H, Chiu H. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
22. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
23. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020 Jun 30;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
24. Kletzel SL, Sood P, Negm A, Heyn PC, Krishnan S, Machtinger J, et al. Effectiveness of brain gaming in older adults with cognitive impairments: a systematic review and meta-analysis. J Am Med Dir Assoc 2021 Nov;22(11):2281-2288.e5 [FREE Full text] [doi: 10.1016/j.jamda.2021.05.022] [Medline: 34146517]
25. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700-b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
26. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
27. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing dataundertaking meta‐analyses. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Sussex, UK: John Wiley & Sons; 2008:243-296.

XSL•FO
RenderX

29. Schunemann H, Oxman A, Vist G, Higgins J, Deeks J, Glasziou P. Chapter 12: Interpreting results and drawing conclusions. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Sussex, UK: John Wiley & Sons; 2008:359-387.
30. Landis JR, Koch GG. An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977 Jun;33(2):363. [doi: 10.2307/2529786]
31. Cavallo M, Hunter EM, van der Hiele K, Angilletta C. Computerized structured cognitive training in patients affected by early-stage Alzheimer's disease is feasible and effective: a randomized controlled study. Arch Clin Neuropsychol 2016 Dec 01;31(8):868-876. [doi: 10.1093/arclin/acw072] [Medline: 27600448]
32. Finn M, McDonald S. Repetition-lag training to improve recollection memory in older people with amnestic mild cognitive impairment. A randomized controlled trial. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2015;22(2):244-258. [doi: 10.1080/13825585.2014.915918] [Medline: 24820545]
33. Yang Y, Kwak YT. Improvement of cognitive function after computer-based cognitive training in early stage of Alzheimer's dementia. Dement Neurocogn Disord 2017 Mar;16(1):7-11 [FREE Full text] [doi: 10.12779/dnd.2017.16.1.7] [Medline: 30906364]
34. Zhuang J, Fang R, Feng X, Xu X, Liu L, Bai Q, et al. The impact of human-computer interaction-based comprehensive training on the cognitive functions of cognitive impairment elderly individuals in a nursing home. J Alzheimers Dis 2013;36(2):245-251. [doi: 10.3233/JAD-130158] [Medline: 23587747]
35. Thapa N, Park HJ, Yang J, Son H, Jang M, Lee J, et al. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. J Clin Med 2020 Apr 29;9(5):1 [FREE Full text] [doi: 10.3390/jcm9051283] [Medline: 32365533]
36. Tarnanas I, Tsolakis A, Tsolaki M. Assessing virtual reality environments as cognitive stimulation method for patients with MCI. In: Brooks A, Brahnam S, Jain L, editors. Technologies of Inclusive Well-Being. Studies in Computational Intelligence, vol 536. Berlin, Heidelberg: Springer; 2014:39-74.
37. Fiatarone Singh MA, Gates N, Saigal N, Wilson GC, Meiklejohn J, Brodaty H, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. J Am Med Dir Assoc 2014 Dec;15(12):873-880. [doi: 10.1016/j.jamda.2014.09.010] [Medline: 25444575]
38. Amjad I, Toor H, Niazi IK, Pervaiz S, Jochumsen M, Shafique M, et al. Xbox 360 Kinect Cognitive Games Improve Slowness, Complexity of EEG, and Cognitive Functions in Subjects with Mild Cognitive Impairment: A Randomized Control Trial. Games Health J 2019 Apr;8(2):144-152. [doi: 10.1089/g4h.2018.0029] [Medline: 30239213]
39. Hagovská M, Takáč P, Dzvoník O. Effect of a combining cognitive and balanced training on the cognitive, postural and functional status of seniors with a mild cognitive deficit in a randomized, controlled trial. Eur J Phys Rehabil Med 2016 Feb;52(1):101-109 [FREE Full text] [Medline: 26325026]
40. van Santen J, Dröes RM, Twisk JWR, Blanson Henkemans OA, van Straten A, Meiland FJM. Effects of exergaming on cognitive and social functioning of people with dementia: a randomized controlled trial. J Am Med Dir Assoc 2020 Dec;21(12):1958-1967.e5 [FREE Full text] [doi: 10.1016/j.jamda.2020.04.018] [Medline: 32651132]
41. Karssemeijer EGA, Aaronson JA, Bossers WJR, Donders R, Olde Rikkert MGM, Kessels RPC. The quest for synergy between physical exercise and cognitive stimulation via exergaming in people with dementia: a randomized controlled trial. Alzheimers Res Ther 2019 Jan 05;11(1):3 [FREE Full text] [doi: 10.1186/s13195-018-0454-z] [Medline: 30611286]
42. Liao Y, Chen I, Hsu W, Tseng H, Wang R. Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: A randomised controlled trial. Ann Phys Rehabil Med 2021 Sep;64(5):101492. [doi: 10.1016/j.rehab.2021.101492] [Medline: 33454398]
43. Flak MM, Hol HR, Hernes SS, Chang L, Engvig A, Bjuland KJ, et al. Adaptive computerized working memory training in patients with mild cognitive impairment. A randomized double-blind active controlled trial. Front Psychol 2019;10:807 [FREE Full text] [doi: 10.3389/fpsyg.2019.00807] [Medline: 31031677]
44. Hyer L, Scott C, Atkinson MM, Mullen CM, Lee A, Johnson A, et al. Cognitive training program to improve working memory in older adults with MCI. Clin Gerontol 2016;39(5):410-427. [doi: 10.1080/07317115.2015.1120257] [Medline: 29471774]
45. Park J, Park J. Does cognition-specific computer training have better clinical outcomes than non-specific computer training? A single-blind, randomized controlled trial. Clin Rehabil 2018 Feb;32(2):213-222. [doi: 10.1177/0269215517719951] [Medline: 28726492]
46. Lee GJ, Bang HJ, Lee KM, Kong HH, Seo HS, Oh M, et al. A comparison of the effects between 2 computerized cognitive training programs, Bettercog and COMCOG, on elderly patients with MCI and mild dementia: A single-blind randomized controlled study. Medicine (Baltimore) 2018 Nov;97(45):e13007 [FREE Full text] [doi: 10.1097/MD.0000000000013007] [Medline: 30407291]
47. Abd-Alrazaq A, Alajlani M, Alhuwail D, Toro CT, Giannicchi A, Ahmed A, et al. The effectiveness and safety of serious games for improving cognitive abilities among elderly people with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Mar 10;10(1):e34592 [FREE Full text] [doi: 10.2196/34592] [Medline: 35266877]

XSL•FO
RenderX

48. Bhattacherjee A. Social Science Research: Principles, Methods, and Practices. Scotts Valley, CA: CreateSpace Independent Publishing Platform; 2012.
49. Mobile Statistics Report, 2021-2025. The Radicati Group, Inc. URL: https://www.radicati.com/wp/wp-content/uploads/2021/Mobile_Statistics_Report,_2021-2025_Executive_Summary.pdf [accessed 2022-07-04]

Abbreviations

AD: Alzheimer disease
CBT: cognitive behavioral therapy
DALYs: Disability-adjusted life years
MCI: mild cognitive impairment
MCID: minimal clinically important difference
MMSE: Mini-Mental State Examination
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: Randomized Controlled Trial
RoB-2: Risk-of-Bias 2
SMD: standardized mean difference
WHO: World Health Organization

Edited by N Zary; submitted 05.01.22; peer-reviewed by F Shuweihdi, L Lamparska; comments to author 06.03.22; revised version received 07.03.22; accepted 16.03.22; published 25.07.22.

Please cite as:
Abd-alrazaq A, Alhuwail D, Ahmed A, Househ M
Effectiveness of Serious Games for Improving Executive Functions Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis
JMIR Serious Games 2022;10(3):e36123
URL: https://games.jmir.org/2022/3/e36123
doi:10.2196/36123
PMID:35877166

©Alaa Abd-alrazaq, Dari Alhuwail, Arfan Ahmed, Mowafa Househ. Originally published in JMIR Serious Games (https://games.jmir.org), 25.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Long-term Effects of Immersive Virtual Reality Reminiscence in People With Dementia: Longitudinal Observational Study

Ling-Chun Huang[1,2], MD, MSc; Yuan-Han Yang[1,2,3,4], PhD

[1]Department of Neurology, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

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

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

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

Corresponding Author:
Yuan-Han Yang, PhD
Department of Neurology
Kaohsiung Municipal Ta-Tung Hospital
Kaohsiung Medical University
68 Jhonghua 3rd Road
Cianjin District
Kaohsiung, 80145
Taiwan
Phone: 886 7 2911101 ext 8913
Fax: 886 7 3162158
Email: endlessyhy@gmail.com

Abstract

Background: Novel nonpharmacological therapies are being developed to prevent cognitive decline and reduce behavioral and psychological symptoms in patients with dementia. Virtual reality (VR) reminiscence was reported to improve anxiety, apathy, and cognitive function immediately after intervention in individuals at residential aged care facilities. However, its effect on elderly patients with dementia and how long this effect could last remain unknown.

Objective: The aim of this paper is to investigate the effect of immersive VR reminiscence in people with dementia both immediately after and 3-6 months after intervention.

Methods: A pilot study was conducted in 2 dementia care units. VR reminiscence therapy sessions were conducted twice per week for a 3-month period. Cognitive function, global status, depressive symptoms, and caregiver burden were assessed before and immediately after VR intervention in 20 participants. Subsequently, 7 participants were reassessed 3-6 months after the VR intervention. Wilcoxon sign-rank test was used for statistical comparisons of the changes.

Results: There were no significant changes in cognitive function, global status, and caregiver burden immediately after the VR intervention, but there was a significant reduction in depressive symptoms (P=.008). Moreover, compared with the cognitive function immediately after VR, it kept declining 3-6 months after.

Conclusions: Immersive VR reminiscence can improve mood and preserve cognitive function in elderly patients with dementia during the period of the intervention. Studies using a control group and comparing the use of VR with traditional forms of reminiscence should be conducted in the future to confirm and expand on these findings.

(JMIR Serious Games 2022;10(3):e36720) doi:10.2196/36720

KEYWORDS

virtual reality; reminiscence; dementia; long-term care

Introduction

Over the past few decades, as the elderly population has rapidly increased, the prevalence of dementia has also continuously increased in high-income countries, and cognitive decline has had a strong impact on society and the economy [1]. Behavioral and psychological symptoms of dementia (BPSD) are frequently noted in patients with dementia, causing distress, reducing the individual's quality of life, exacerbating cognitive and functional impairment, and increasing caregiver stress [2]. Although

XSL•FO
RenderX

pharmacological interventions are recommended for the treatment of cognitive impairment and BPSD in Alzheimer disease (AD) and other types of dementia, polypharmacy may cause side effects in the elderly population. Novel and effective nonpharmacological therapies are therefore required.

Virtual reality (VR) is a computer-generated environment that has been developed in recent years. Using headsets, it can provide a fully immersive, highly engaging, and realistic experience. VR is often used for recreational purposes, but it is increasingly being used in medical training, rehabilitation, and therapy [3-5]. Current research suggests that immersive VR is safe and well tolerated, and it can promote engagement and provide an enjoyable experience for people with dementia [6-8].

Reminiscence therapy, one of the most popular psychosocial interventions, involves the discussion of past activities, events, and experiences, usually with the aid of tangible prompts, such as photographs or music [9]. Reminiscence therapy has been shown to have some positive effects for people with dementia by increasing their quality of life, cognition, communication, and mood, although the noted effects were small [9]. Reminiscence therapy using immersive VR may be more effective as it would be more realistic than traditional reminiscence therapy and could lead to increased engagement. In previous studies, immersive VR reminiscence therapy was shown to reduce anxiety and apathy while improving semantic verbal fluency, immediately after a short intervention program in elderly individuals residing in aged care facilities with different cognitive statuses [10,11]. However, the long-term effects in people with dementia remain unknown.

Therefore, we conducted a study to investigate the immediate and 3-6–month effects of immersive VR reminiscence therapy in people with dementia.

Methods

Recruitment

The participants were recruited from 2 dementia care units in Kaohsiung city, Taiwan. Potentially suitable participants were identified by the staff at the dementia care units based on the inclusion and exclusion criteria. The inclusion criteria were (1) those diagnosed with all-cause dementia by experienced physicians, based on the National Institute on Aging and Alzheimer's Association diagnostic criteria [12]; and (2) attended the dementia care units between July 2020 and March 2021. Participants were excluded if (1) they had poor recognition of the VR images even after adjusting the mounting position of the headset; and (2) their cognitive function was too low or their BPSD was too severe, making assessment difficult.

Ethics Approval

The participants and their relatives were informed of the details of the study prior to their inclusion, and appropriate written informed consent was obtained. The Kaohsiung Medical University Hospital Institutional Review Board (KMUHIRB-SV(II)-20190049) approved the study protocol.

VR Apparatus

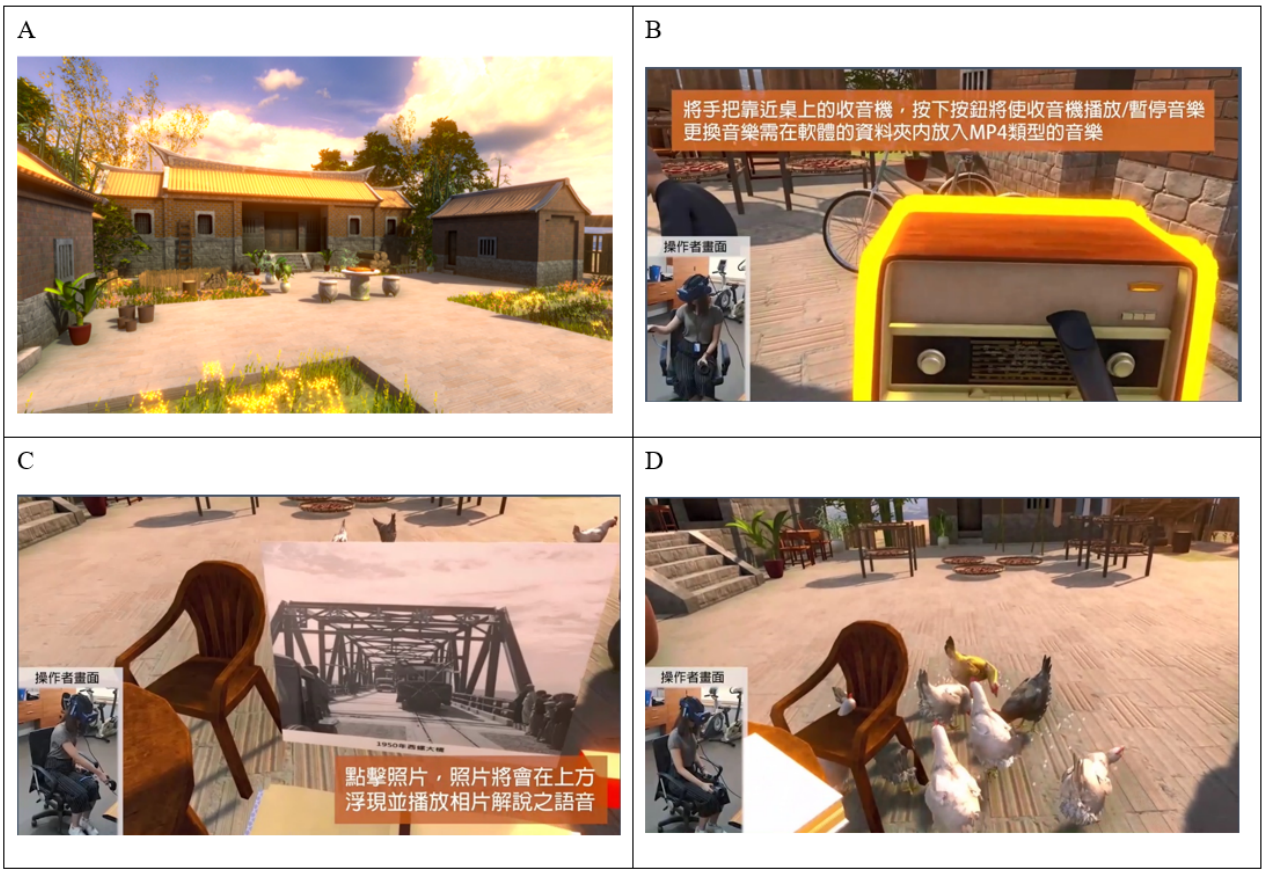
The VIVE Pro VR head-mounted display (HMD) was used to deliver the immersive VR experience to the participants. It is a stand-alone HMD providing stereoscopic vision at a resolution of 2880 x 1600 per eye with a 90 Hz refresh rate. Two controllers were paired with the HMD to enable the participants to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

Preparation of VR Content

Computer graphics VR images were created based on a historical type of residence that was commonly found throughout Taiwan in 1960-1980 AD (Figure 1). Photographs, narration, and music of personal past significance to the participants were provided by the family or caregivers on an individual basis. These materials were transformed into digital forms and integrated into the VR content. The participants could use the controllers to turn on the radio to play the music and to look through the photo album and browse the photographs with a voice narrating the past meaningful situation and events about the photographs. They could also use the controller to hold rice to feed chickens, which was traditional in many old villages.

XSL•FO
RenderX

Figure 1. Scenes from the computer generated virtual reality reminiscence. Computer graphics images based on (A) a historical type of residence that was commonly found throughout Taiwan between 1960-1980 AD. (B) The participants could use the controllers to turn on the radio to play music and (C) to open the photo album to browse the photographs with a voice narrating. (D) They could also use the controllers to hold rice to feed chickens, which was traditional in older villages.



Implementation

The VR intervention was administered twice per week over a period of 3 months. The participants viewed and interacted with the VR content for approximately 10 to 12 minutes each time. The HMD images were mirrored to a laptop to enable the researcher to see what the participants were viewing and interacting with. The participants remained seated during the VR intervention to reduce motion sickness when using the HMD and to also minimize the risk of falls in the elderly. During VR intervention, there was a conversation between the researcher and the participant regarding the content being viewed. The participants were provided with the controllers to enable them to interact with the VR space, and the researcher was beside them to assist and guide them when needed.

Assessment of Questionnaire and Scores

Before and after the VR intervention, the participants were assessed on their cognitive function using the Cognitive Abilities Screening Instrument (CASI), the Mini-Mental State Examination (MMSE), the global status by Clinical Dementia Rating (CDR), and the depressive symptoms by Center for Epidemiological Studies Depression (CESD). Caregiver burden was evaluated using the Zarit Caregiver Burden Interview (ZBI).

Cognitive Abilities Screening Instrument

The CASI was used to objectively assess overall cognitive function. The CASI has a score range of 0 to 100 and provides a quantitative assessment of attention, concentration, orientation, short-term memory, long-term memory, language abilities, visual construction, list-generating fluency, abstraction, and judgment [13]. Pilot testing has demonstrated its usefulness in monitoring disease progression and in providing profiles of cognitive impairment [13].

Mini-Mental State Examination

Cognitive impairment was also assessed using the MMSE, which assesses orientation, immediate recall, calculation or attention, delayed recall, naming, repetition, 3-stage command, reading, writing, and constructional praxis. The total score is 30, and a lower score indicates more severe cognitive impairment [14].

Clinical Dementia Rating

The global status of dementia was defined according to the CDR. The CDR is a widely used clinical staging instrument for characterizing the manifestation and severity of dementia. It is generated from a semistructured interview with the patient and a knowledgeable informant evaluating six cognitive domains (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care). The overall CDR is derived by providing ratings for each of the six domains, where a score of 0 indicates normal; a score of 0.5 signifies uncertain or very mild impairment; and a score of 1, 2, or 3 corresponds to mild, moderate, or severe impairment

XSL•FO
RenderX

[15]. The scores for the six domains can then be added together (CDR Sum of Boxes [CDR-SB]).

Center for Epidemiological Studies Depression

Depressive symptoms were assessed using the CESD, a 20-item self-administered questionnaire. On a 4-point Likert scale ranging from 0 (none) to 3 (5 days or more), the participants were asked to rate the frequency with which each item occurred every week. Higher scores indicate more severe depressive symptoms [16].

Zarit Caregiver Burden Interview

The ZBI is a well-known self-reporting measure of perceived burden among caregivers. The instrument measures the caregiver's emotion, psychological health, well-being, social and family life, finances, and degree of control over one's life. The version used contains 22 items, and each item on the questionnaire is a statement that the caregiver is asked to endorse on a 5-point Likert scale (0: never; 1: rarely; 2: sometimes; 3: quite frequently; and 4: nearly always) [17]. Higher scores indicate higher caregiver burden.

Statistical Analysis

Data were collected from July 2020 to March 2021. In the results, the data are presented as mean (SD) or as a proportion. Comparisons of scores before and immediately after VR intervention, as well as immediately after and 3-6 months after VR intervention, were performed using the Wilcoxon sign-rank test. Analyses were performed using SPSS 26.0 (IBM Corporation). A 2-tailed *P* value of <.05 was considered to indicate a statistically significant difference.

Results

Baseline Demographic Characteristics, Cognitive Function, and Global Status

A total of 25 individuals were enrolled in the study at the beginning; 2 (8%) were excluded because they left the dementia care units, and 3 (12%) were excluded as they were unable to receive a postintervention evaluation due to physical problems. The baseline demographic characteristics, cognitive function, and global status of the 20 participants who completed the study are shown in Table 1. The participants' mean age was 79.0 (SD 7.8) years; 45% (9/20) were male and 55% (11/20) were female. The mean education level was 10.8 (SD 4.3) years. The mean CASI score was 55.5 (SD 15.6), and the mean MMSE score was 15.4 (SD 5.5). According to CDR, 2/20 (10%) participants had very mild dementia, 15/20 (75%) participants had mild dementia, and 3/20 (15%) participants had moderate dementia.

Table 1. Baseline demographic characteristics, cognitive function, and global status of the participants.

Characteristics	Values (N=20)
Age (years), mean (SD)	79.0 (7.8)
Sex, n (%)	
Male	9 (45)
Female	11 (45)
Education (years), mean (SD)	10.8 (4.3)
CASI[a], mean (SD)	55.5 (15.6)
MMSE[b], mean (SD)	15.4 (5.5)
CDR[c], n (%)	
0.5	2 (10)
1	15 (75)
2	3 (15)
CDR sum of boxes, mean (SD)	7.3 (3.0)

[a]CASI: Cognitive Assessment Screening Instrument.

[b]MMSE: Mini-Mental State Examination.

[c]CDR: Clinical Dementia Rating.

The Immediate Effect of VR Reminiscence

The immediate effects of immersive VR on cognition, global status, depressive symptoms, and caregiver burden are shown in Table 2. There were no significant differences in the MMSE, CASI and its subdomains, CDR-SB, and ZBI scores before and immediately after VR intervention, while the scores for CESD significantly decreased from 6.15 (SD 5.73) to 3.15 (SD 4.26; *P*=.008).

XSL•FO
RenderX

Table 2. The immediate effect of virtual reality (VR) reminiscence on cognition, global status, depressive symptoms, and caregiver burden. The comparison of scores before and after VR were evaluated by Wilcoxon sing-rank test.

Characteristics	Before VR, mean (SD)	Immediately after VR, mean (SD)	*P* value
MMSE[a]	15.40 (5.47)	14.95 (5.13)	.48
CASI[b]	55.47 (15.64)	54.46 (16.64)	.22
Remote memory (0-10)	8.75 (1.80)	8.25 (2.51)	.13
Orientation (0-18)	6.65 (3.86)	6.85 (3.68)	.52
Attention (0-8)	5.85 (1.43)	5.65 (1.63)	.54
Concentration (0-10)	5.85 (2.91)	5.95 (3.12)	.77
Recent memory (0-12)	1.59 (1.55)	1.66 (1.88)	.46
Fluency (0-10)	3.30 (2.00)	3.40 (2.42)	.95
Language (0-10)	7.58 (2.39)	7.45 (1.99)	.70
Abstraction (0-6)	3.00 (1.69)	2.80 (1.32)	.30
Judgment (0-6)	4.65 (1.27)	4.75 (1.21)	.74
Visual construction (0-10)	8.25 (2.75)	7.70 (3.33)	.10
CDR[c] sum of boxes	7.26 (3.00)	7.33 (2.43)	.64
CESD[d]	6.15 (5.73)	3.15 (4.26)	*.008* [e]
ZBI[f]	34.65 (15.98)	31.20 (14.05)	.14

[a]MMSE: Mini-Mental State Examination.

[b]CASI: Cognitive Assessment Screening Instrument.

[c]CDR: Clinical Dementia Rating.

[d]CESD: Center for Epidemiological Studies Depression.

[e]Significant value is shown in italics.

[f]ZBI: Zarit Caregiver Burden Interview.

The Long-term Effect of VR Reminiscence

Among the 20 participants, 7 (35%) were followed up 3-6 months after the VR intervention. No significant changes in MMSE, CASI and CDR-SB were noted before and immediately after VR in these 7 participants; however, the CASI score significantly decreased 3-6 months after VR, compared to immediately after VR (52.14, SD 15.71 vs 57.50, SD 12.40; *P*=.03; Tables 3 and 4; Figure 2).

Table 3. Cognition and global status before and immediately after virtual reality (VR) reminiscence in 7 participants. The comparison of scores were evaluated by Wilcoxon sign-rank test.

Characteristics	Before VR, mean (SD)	Immediately after VR, mean (SD)	*P* value
MMSE[a]	15.57 (4.76)	16.29 (4.07)	.67
CASI[b]	58.80 (12.48)	57.50 (12.40)	.50
CDR[c] sum of boxes	6.50 (1.61)	6.86 (1.28)	.10

[a]MMSE: Mini-Mental State Examination.

[b]CASI: Cognitive Assessment Screening Instrument.

[c]CRD: Clinical Dementia Rating.

XSL•FO
RenderX

Table 4. Cognition and global status immediately after and 3-6 months after virtual reality (VR) reminiscence in 7 participants. The comparison of scores were evaluated by Wilcoxon sign-rank test.

Characteristics	Immediately after VR, mean (SD)	3-6 months after VR, mean (SD)	*P* value
MMSE[a]	16.29 (4.07)	16.00 (5.23)	.75
CASI[b]	57.50 (12.40)	52.14 (15.71)	*.03* [c]
CDR[d] sum of boxes	6.86 (1.28)	7.93 (2.86)	.31

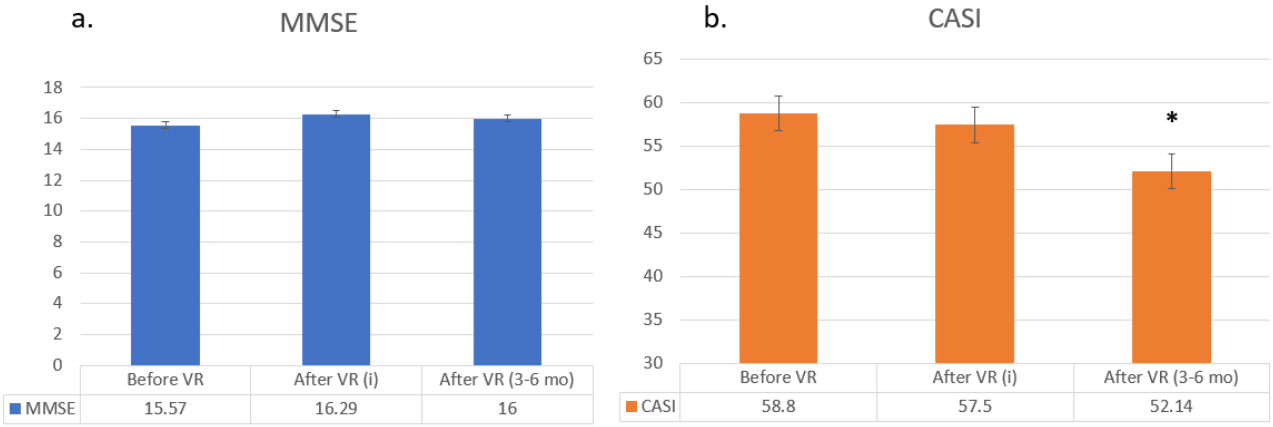
[a]MMSE: Mini-Mental State Examination.

[b]CASI: Cognitive Assessment Screening Instrument.

[c]Significant value shown in italics.

[d]CDR: Clinical Dementia Rating.

Figure 2. Mini-Mental State Examination (MMSE) and Cognitive Abilities Screening Instrument (CASI) scores before, immediately after, and 3-6 months after virtual reality (VR) reminiscence in 7 participants. Data are shown as mean (SD) for quantitative variables. A comparison of a) MMSE and b) CASI scores before and immediately after VR as well as immediately after VR and 3-6 months after VR were evaluated by Wilcoxon sign-rank test (*P*<.05 is statistically significant); i: immediately.



Discussion

Principal Results

In this paper, we investigated the potential effects of immersive VR reminiscence therapy in people with dementia, as well as studying how long the effects could last. Although there were no obvious changes in cognition, global status, and caregiver burden after the VR intervention, the depressive symptoms improved significantly after VR therapy. Compared to the CASI scores immediately after VR, the scores after 3-6 months were significantly decreased. In other words, immersive VR reminiscence may improve mood and preserve cognitive function in elderly patients with dementia during the period of intervention. To the best of our knowledge, this is the first study to explore not only the possible effects of immersive VR reminiscence but also the effect period in people with dementia.

Comparison With Prior Work

Although the use of VR is being developed in many different fields, evidence regarding reminiscence interventions with immersive VR in patients with dementia remains limited. Niki et al [10] found that immersive VR reminiscence could reduce anxiety in the elderly living in a nursing home, without causing serious side effects. However, most of their participants had preserved cognitive function at baseline. Saredakis et al [11] recruited 17 older adults residing in an aged care facility for tailored VR reminiscence and found that it could improve apathy and semantic verbal fluency immediately after the intervention. However, more than half of the participants in that study had no or minimal cognitive impairment. Contrary to these previous studies, we recruited participants who were attending dementia care units to explore the effect of immersive VR reminiscence in individuals with dementia.

People with dementia have a high incidence of depression, which reduces the quality of life for both patients and caregivers and is associated with increased costs and reduced cognition [18]. Although reminiscence therapy may have some benefits in reducing depression in people with dementia, its effectiveness should be tested further [19]. In a multicenter randomized controlled trial, reminiscence therapy failed to improve depressive symptoms in older adults with dementia [20,21], whereas in our study, VR reminiscence significantly reduced depressive symptoms. As an interesting and enjoyable tool, VR intervention may be more effective than traditional reminiscence in improving mood. Further studies comparing VR reminiscence with traditional forms of reminiscence should be conducted in the future.

Based on previous studies, it was not known how long the effects of VR could last. Therefore, we reassessed 7 of the study participants 3-6 months after the VR intervention, and we found that their cognitive function kept declining after discontinuing the intervention. Because more than half of the etiology of

XSL•FO
RenderX

dementia is AD, a degenerative disease with an average MMSE decrease of 1.15 points and a CASI decrease of 4.27 points per year [22], the fact that there was no obvious change in cognition during the VR intervention implies that the VR reminiscence therapy may maintain cognition or reduce cognitive decline in people with dementia. Further studies with maintained VR intervention should be conducted to confirm this.

It is reasonable that enough sessions of VR are required to obtain therapeutic effects. In a previous study promoting VR reminiscence in people with dementia, there were no significant changes in the psychological and behavioral symptoms and in the quality of life after a short course of intervention, even though the caregivers assessed the experience as potentially beneficial for most participants [23]. Park et al [24] suggested that reminiscence therapy of more than 8 sessions might be required to obtain any therapeutic effects. Therefore, in our study, the VR intervention was administered twice per week with a period of 3 months.

Our study design was different from that of most previous studies, and our participants could use the controllers to interact with the virtual environment. Interactive VR has been found to increase the sense of presence and to have a positive effect on the immersive experience [25]. Recent research has found that the use of VR in providing interactions may be an alternative way of delivering stimulation to people with dementia who do not participate in other lifestyle activities [26], and interactions can be recorded during the VR experience. In future studies, recording of the participants' limb movement while they use the controllers could enable analysis of the speed and accuracy of their movements.

Limitations

There were some limitations to this study. First, there was no control group or a group receiving traditional reminiscence therapy; therefore, the results of this study must be carefully interpreted. Second, the number of subjects was small, especially in the reassessed cases. Studies with larger sample sizes should be conducted in the future to validate our findings. Third, the staff at the dementia care units identified the potentially suitable participants for the study; therefore, the effects may not be applicable to all cases of dementia. Fourth, our VR reminiscence was not totally personalized because computer graphics take a lot of time to create and are expensive. Some VR images were different from those in the subjects' own memories, and using personalized content has been demonstrated to be more effective than using generic content [27]. To combat this, we collected personal photographs and music from the participants' past experiences, which could be used to tailor the VR experience. Despite the reported limitations, this study is a pilot study to explore not only the potential effect of immersive VR reminiscence but also how long the effect existed in people with dementia. We also objectively evaluated the possible effects on caregiver burden. During the COVID-19 pandemic, digital therapeutics such as VR are important for facilitating remote health care and for reducing the risk of cluster infections.

Conclusions

Our study found that immersive VR reminiscence may improve mood and preserve cognitive function in elderly patients with dementia during the period of the intervention. Studies using a control group and comparing the use of VR with traditional forms of reminiscence should be conducted in the future to confirm and develop these findings.

Acknowledgments

This study was supported by grants from National Health Research Institutes Grant (PH-110-GP-08, PH-111-GP-08), Ministry of Science and Technology (MOST 108-2745-8-037-011-), Kaohsiung Medical University Research Center (KMU-TC110B03), and Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan (KMTTH-110-027). We also thank HamaStar Technology Co, Ltd for their assistance.

Conflicts of Interest

None declared.

References

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007 Jul;3(3):186-191. [doi: 10.1016/j.jalz.2007.04.381] [Medline: 19595937]
2. Black W, Almeida OP. A systematic review of the association between the Behavioral and Psychological Symptoms of Dementia and burden of care. Int Psychogeriatr 2004 Sep;16(3):295-315. [doi: 10.1017/s1041610204000468] [Medline: 15559754]
3. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual Reality and Medical Inpatients: A Systematic Review of Randomized, Controlled Trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
4. Rose T, Nam CS, Chen KB. Immersion of virtual reality for rehabilitation - Review. Appl Ergon 2018 May;69:153-161. [doi: 10.1016/j.apergo.2018.01.009] [Medline: 29477323]
5. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]

XSL•FO
RenderX

6. Kim O, Pang Y, Kim J. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]
7. Brimelow RE, Dawe B, Dissanayaka N. Preliminary Research: Virtual Reality in Residential Aged Care to Reduce Apathy and Improve Mood. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):165-170. [doi: 10.1089/cyber.2019.0286] [Medline: 31829729]
8. Manera V, Chapoulie E, Bourgeois J, Guerchouche R, David R, Ondrej J, et al. A Feasibility Study with Image-Based Rendered Virtual Reality in Patients with Mild Cognitive Impairment and Dementia. PLoS One 2016 Mar 18;11(3):e0151487 [FREE Full text] [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
9. Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 01;3:CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
10. Niki K, Yahara M, Inagaki M, Takahashi N, Watanabe A, Okuda T, et al. Immersive Virtual Reality Reminiscence Reduces Anxiety in the Oldest-Old Without Causing Serious Side Effects: A Single-Center, Pilot, and Randomized Crossover Study. Front Hum Neurosci 2020;14:598161 [FREE Full text] [doi: 10.3389/fnhum.2020.598161] [Medline: 33536887]
11. Saredakis D, Keage HA, Corlis M, Loetscher T. Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed Methods Study. J Med Internet Res 2020 Jun 26;22(6):e17632 [FREE Full text] [doi: 10.2196/17632] [Medline: 32469314]
12. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
13. Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. Int Psychogeriatr 1994;6(1):45-58; discussion 62. [doi: 10.1017/s1041610294001602] [Medline: 8054493]
14. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
15. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
16. Cheng C, Yen C, Ko C, Yen J. Factor structure of the Center for Epidemiologic Studies Depression Scale in Taiwanese adolescents. Compr Psychiatry 2012 Apr;53(3):299-307. [doi: 10.1016/j.comppsych.2011.04.056] [Medline: 21621755]
17. Hébert R, Bravo G, Préville M. Reliability, Validity and Reference Values of the Zarit Burden Interview for Assessing Informal Caregivers of Community-Dwelling Older Persons with Dementia. Can. J. Aging 2010 Nov 29;19(4):494-507. [doi: 10.1017/S0714980800012484]
18. Enache D, Winblad B, Aarsland D. Depression in dementia. Current Opinion in Psychiatry 2011;24(6):461-472. [doi: 10.1097/yco.0b013e32834bb9d4]
19. Thomas JM, Sezgin D. Effectiveness of reminiscence therapy in reducing agitation and depression and improving quality of life and cognition in long-term care residents with dementia: A systematic review and meta-analysis. Geriatr Nurs 2021;42(6):1497-1506. [doi: 10.1016/j.gerinurse.2021.10.014] [Medline: 34735996]
20. Justo-Henriques SI, Pérez-Sáez E, Alves Apóstolo JL. Multicentre randomised controlled trial about the effect of individual reminiscence therapy in older adults with neurocognitive disorders. Int J Geriatr Psychiatry 2021 May;36(5):704-712. [doi: 10.1002/gps.5469] [Medline: 33176394]
21. Pérez-Sáez E, Justo-Henriques SI, Alves Apóstolo JL. Multicenter randomized controlled trial of the effects of individual reminiscence therapy on cognition, depression and quality of life: Analysis of a sample of older adults with Alzheimer's disease and vascular dementia. Clin Neuropsychol 2021 Jan 19:1-22. [doi: 10.1080/13854046.2021.1871962] [Medline: 33467972]
22. Yang Y, Wu M, Chou P, Su H, Lin S, Sung P. Longitudinal Neuropsychological Outcome in Taiwanese Alzheimer's Disease Patients Treated with Medication. Curr Alzheimer Res 2018 Mar 14;15(5):474-481. [doi: 10.2174/1567205014666171010112518] [Medline: 29032750]
23. Coelho T, Marques C, Moreira D, Soares M, Portugal P, Marques A, et al. Promoting Reminiscences with Virtual Reality Headsets: A Pilot Study with People with Dementia. Int J Environ Res Public Health 2020 Dec 12;17(24):9301 [FREE Full text] [doi: 10.3390/ijerph17249301] [Medline: 33322679]
24. Park K, Lee S, Yang J, Song T, Hong GS. A systematic review and meta-analysis on the effect of reminiscence therapy for people with dementia. Int. Psychogeriatr 2019 Feb 04;31(11):1581-1597. [doi: 10.1017/s1041610218002168]
25. Yeo NL, White MP, Alcock I, Garside R, Dean SG, Smalley AJ, et al. What is the best way of delivering virtual nature for improving mood? An experimental comparison of high definition TV, 360° video, and computer generated virtual reality. J Environ Psychol 2020 Dec;72:101500 [FREE Full text] [doi: 10.1016/j.jenvp.2020.101500] [Medline: 33390641]
26. Baker S, Waycott J, Robertson E, Carrasco R, Neves B, Hampson R, et al. Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. Information Processing & Management 2020 May;57(3):102105 [FREE Full text] [doi: 10.1016/j.ipm.2019.102105]

XSL•FO
RenderX

27. Yasuda K, Kuwabara K, Kuwahara N, Abe S, Tetsutani N. Effectiveness of personalised reminiscence photo videos for individuals with dementia. Neuropsychol Rehabil 2009 Aug;19(4):603-619. [doi: 10.1080/09602010802586216] [Medline: 19235628]

Abbreviations

AD: Alzheimer disease
BPSD: behavioral and psychological symptoms of dementia
CASI: Cognitive Abilities Screening Instrument
CDR-SB: Clinical Dementia Rating-Sum of Boxes
CDR: Clinical Dementia Rating
CESD: Center for Epidemiological Studies Depression
HMD: head-mounted display
MMSE: Mini-Mental State Examination
VR: virtual reality
ZBI: Zarit Caregiver Burden Interview

Edited by N Zary; submitted 21.01.22; peer-reviewed by YT Lin, J Guendouzi; comments to author 12.06.22; revised version received 14.06.22; accepted 09.07.22; published 25.07.22.

Please cite as:
Huang LC, Yang YH
The Long-term Effects of Immersive Virtual Reality Reminiscence in People With Dementia: Longitudinal Observational Study
JMIR Serious Games 2022;10(3):e36720
URL: https://games.jmir.org/2022/3/e36720
doi:10.2196/36720
PMID:35877169

©Ling-Chun Huang, Yuan-Han Yang. Originally published in JMIR Serious Games (https://games.jmir.org), 25.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tutorial

Developing Serious Video Games to Treat Attention Deficit Hyperactivity Disorder: Tutorial Guide

Aarón Súja r[1,2*], PhD; Marina Martín-Moratinos[1,3*], MSc; María Rodrigo-Yanguas[1,3], PhD; Marcos Bella-Fernández[1,4,5], MSc; Carlos González-Tardón[6], PhD; David Delgado-Gómez[7], PhD; Hilario Blasco-Fontecilla[1,3,8,9], MD, PhD

[1]Department of Psychiatry, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain
[2]Department of Computer Engineering, Universidad Rey Juan Carlos, Madrid, Spain
[3]Faculty of Medicine, Universidad Autónoma de Madrid, Madrid, Spain
[4]Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain
[5]Department of Psychology, Universidad Pontificia de Comillas, Madrid, Spain
[6]Tecnocampus, Universitat Pompeu Fabra, Mataró, Spain
[7]Department of Statistics, Universidad Carlos III, Leganés, Spain
[8]ITA Mental Health, Madrid, Spain
[9]Centro de Investigación Biomédica en Red Salud Mental, Madrid, Spain
[*]these authors contributed equally

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Department of Psychiatry
Hospital Universitario Puerta de Hierro Majadahonda
1 C Joaquín Rodrigo
Majadahonda, 28222
Spain
Phone: 34 911916019
Email: hmblasco@yahoo.es

Abstract

Video game–based therapeutic interventions have demonstrated some effectiveness in decreasing the symptoms of attention deficit hyperactivity disorder (ADHD). Compared with more traditional strategies within the multimodal treatment of ADHD, video games have certain advantages such as being comfortable, flexible, and cost-efficient. However, establishing the most appropriate type(s) of video games that should be used for this treatment remains a matter of debate, including the commercial existing video games or serious video games that are specifically constructed to target specific disorders. This guide represents a starting point for developing serious video games aimed at treating ADHD. We summarize the key points that need to be addressed to generate an effective and motivating game-based treatment. Following recommendations from the literature to create game-based treatments, we describe the development stages of a serious video game for treating ADHD. Game design should consider the interests of future users; game mechanics should be based on cognitive exercises; and therapeutic mechanisms must include the control of difficulty, engagement, motivation, time constraints, and reinforcement. To elaborate upon this guide, we performed a narrative review focused on the use of video games for the treatment of ADHD, and were inspired by our own experience during the development of the game “The Secret Trail of Moon.”

(JMIR Serious Games 2022;10(3):e33884) doi:10.2196/33884

KEYWORDS

serious video games; ADHD; treatment; video games; cognitive; cognitive disorder; games

Introduction

Attention deficit hyperactivity disorder (ADHD) is the most frequent neurodevelopmental disorder with a worldwide prevalence ranging between 4% and 8% [1]. Core ADHD symptoms (inattention, hyperactivity, and impulsivity) negatively impact social, emotional, and/or cognitive domains compromising two or more areas of daily life [2]. The prognosis is clouded by comorbidities such as executive dysfunction [3] or emotional dysregulation, among others [4]. If undetected or

XSL•FO
RenderX

untreated, people with ADHD have a higher probability of accidents, school dropout, addictions, and mortality [5].

According to guidelines, the treatment of choice for ADHD is multimodal [6]. Low adherence to medication is common [7], which is explained by several factors: adverse effects, parent preferences, misuse, and/or omissions [8]. Cognitive behavioral therapy (CBT) is the most effective psychological treatment for ADHD [6]. Although CBT has shown effectiveness in adolescents with ADHD [9], psychotherapies are not usually implemented within public health systems [10]. Furthermore, lack of motivation and engagement are core characteristics of ADHD [4]. Accordingly, any intervention aimed at rehabilitating ADHD should be focused on fun. New technologies, particularly video games, have the potential to fulfill this goal. Indeed, Sonuga-Barke [11] affirms that the playful elements of video games help to maintain motivation for engaging in therapy.

According to the Interactive Software Federation Europe [12], there are currently more than 24 million game players in the age group of 6-15 years across the European market. People with ADHD are more prone to use video games, being up to three times more likely to develop an addiction [13]. However, appropriate use of video games may complement multimodal treatment [14]. Indeed, there is preliminary evidence about the potential use of serious video games to treat ADHD [15-17]. Furthermore, there are guidelines for the creation of health-based games. For example, Duncan et al [18] proposed useful tools (Playbook) for multidisciplinary teams devoted to the development of serious games–based therapies. Kinross [19] proposed a methodology for creating health games inspired by current medical intervention methodologies. Lastly, Baranowski [20] highlights the minimum descriptions that must be considered when developing a new game for health.

The aim of this Tutorial is to provide a basic guide for developers of serious video games targeting ADHD by considering the particularities (eg, lack of adherence, addiction) of this disorder. We offer a roadmap for companies, programmers, and researchers interested in this field. We take advantage of our own experience in developing the serious game "The Secret Trail of Moon" (TSTM) [21-23], a serious video game designed to treat ADHD, which has been tested in a randomized clinical trial (ClinicalTrials.gov NCT04355065).

Benefits of Video Games

Video games offer intrinsic benefits that can be exploited for other purposes. Some studies stressed that video games could have beneficial cognitive effects on attention and visuospatial abilities in adults [24]; can improve attention, effort, and motivation [25]; or can be effective on top-down attentional control, task-switching, processing speed, and increased time perception [26]. Other studies outline the motivational capacity of video games on emotional and social skills [27].

Video games created with nongame objectives, which are known as "serious video games," have been developed for different activities such as training, education, or evaluation in nonludic environments [28]. Serious video game activities are motivationally challenging while simultaneously offering users a fun learning experience. Their narratives and technical resources can increase the level of engagement of the user to pursue a specific objective [29]. Levels of engagement are linked to positive emotions produced by effort and overcoming obstacles, which are essential aspects in turning a video game into a tool [30].

Video game–based therapeutic interventions have proven to be useful in health care environments. As previously mentioned, children with ADHD have low intrinsic motivation and poor internal language, which make it difficult for them to complete tasks, leading to procrastination. This low intrinsic motivation causes them to become bored earlier, relying more on an external stimulus or more engaging tasks [4]. Therapeutic interventions based on gamification and video games can be a good strategy for overcoming this problem. In fact, video game–based therapeutic tools appear to be useful and effective in the treatment of ADHD. Engagement rates with video game–based interventions are generally high, with low rates of dropouts (see [15,17] for reviews).

IDEAL-Games Framework

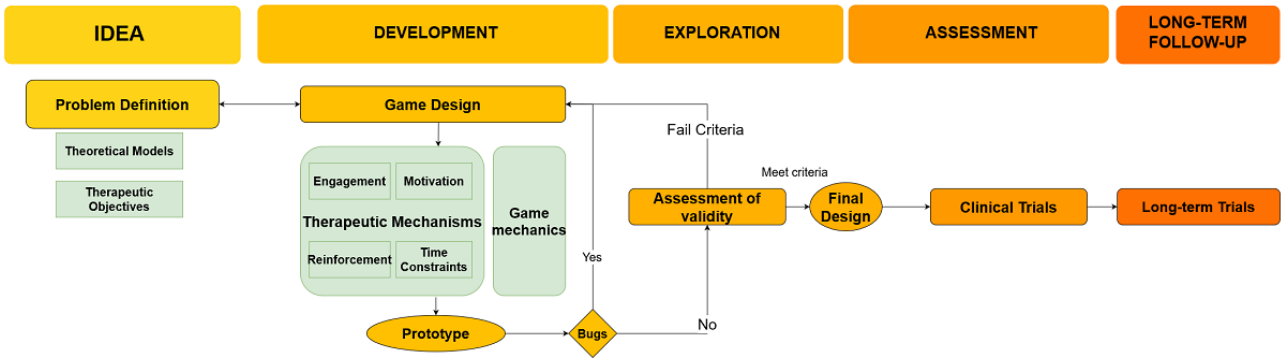
Overview

Taking into account the references on the creation of health-based games mentioned in the Introduction, we have adapted Kinross' [19] proposal and the terminology of Duncan et al [18] and Baranowski [20]. To create a successful therapeutic game, three issues need to be addressed [31]: (1) focus on therapeutic objectives, (2) use game design principles, and (3) incorporate patients' opinions to adjust to their interests. Figure 1 shows the IDEAL-Games proposal. The IDEAL-Games framework describes the video game development process: (1) Idea, (2) Development, (3) Exploration, (4) Evaluation of the effectiveness, and (5) Long-term assessment to understand the effects over time [19].

Throughout the process, the collaboration of a multidisciplinary team is essential to balance clinical and technical components [31]. Duncan et al [18] remarked that clinicians and researchers master the mechanisms of health behavior change but know much less about the creation of an optimal player experience compared with game developers. Therefore, it is important to "create" a common language and for knowledge to be condensed in a Game Playbook [18] or game design document, eliminating jargon and creating a document that can be easily understood by all members of the team. In the case of TSTM [21], the consortium includes: (1) health care professionals, (2) a serious video game consultant, and (3) a game development company.

XSL•FO
RenderX

Figure 1. IDEAL-Games framework [22] adapted to create a general guide for game-based attention deficit hyperactivity disorder treatment.

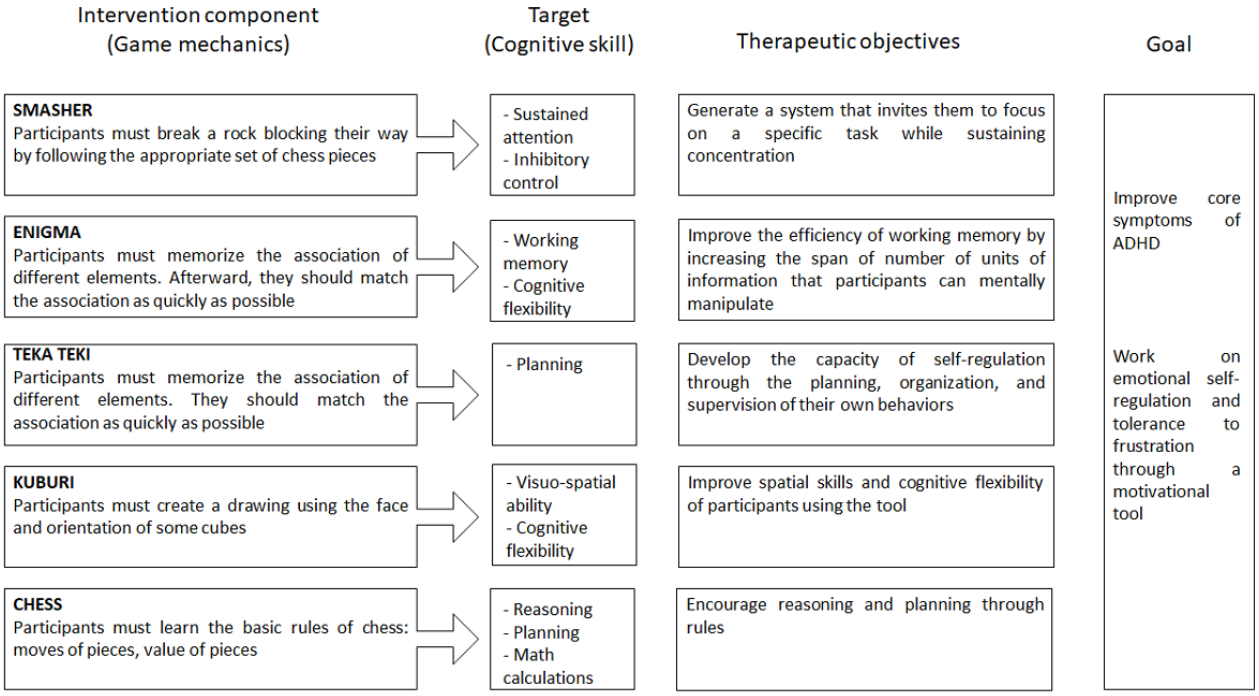


Idea

In the video game industry, a preproduction phase is common. The idea starts by detecting an unmet need. The planning of a project is key to obtain a quality product that is developed within the estimated time and cost. In the case of game-based treatment, the team should stress the main treatment goal, which is based on the literature and the theoretical models chosen. Problem definition determines the final goal. In our example, TSTM [21] was based on the models proposed by Russell Barkley [32] and Thomas Brown [33]. See Rodrigo-Yanguas et al [21,22] for a description of the general characteristics of TSTM.

The idea stage is similar to the preproduction of a video game and refers to the process where a multidisciplinary team identifies the therapeutic goals of the video game. This stage provides a clear overview of what the team aims to achieve during the game-play experience and the first drafts of the game that will be developed in the next stage of the process. They should identify the targeted variables aligned with the therapeutic objectives and the strategies that will be followed to achieve the desired outcome. In the case of TSTM, developing these game mechanics involved hours of discussion within the team to find a balance between psychology and technology: the developers wanted a fun game, while the mental health professionals required internal validity of the game based on psychological constructs. Figure 2 displays the logic model built for TSTM.

Figure 2. Logic model built for The Secret Trail of Moon (based on Duncan et al [18]). ADHD: attention deficit hyperactivity disorder.



The use of new technologies may enhance ecological validity through incorporating complexity, novelty, and diversity in the process of developing the video game [34]. However, choosing an innovative technology represents a challenge. The choice will determine both the game and therapeutics mechanics, and affects the entire developmental stage. Developers must consider the barriers of entry, licensing costs, previous experience, and the scalability of the platform chosen. On the patient side, the developer team must consider the commonly used platforms by the general public, the price, and their commercial availability. The inclusion of these factors can enhance the patient's motivation, isolate the patient from outside distractions, or complement the development of other qualities, which can in turn increase the effectiveness of the treatment. For instance,

XSL•FO
RenderX

TSTM uses Playstation VR [22], Benzing and Schmidt [35] used Microsoft Kinect to develop an exergame, and Lim et al [36] developed an attention training game where children play via the signals detected by electroencephalogram electrodes.

Virtual reality (VR) has been a major challenge in the development of TSTM. An example of this is the Kitsune game [21], which was programmed but could not be finally implemented for two reasons: a failed design (Kitsune failed to work on inhibitory control) and motion sickness (due to an interface very close to the player's vision and being a game in motion). Therefore, we recommend choosing the technology with caution and establishing a good design.

Development

User-Centered Design

In this stage, the multidisciplinary team builds the game's prototype, considering the game components highlighted by Kinross [19]: game mechanics balanced with elements of game design to improve engagement, motivation, and reinforcement.

Since video games intrinsically have a motivational design [29], a serious video game should be similarly designed. In line with the IDEAL-Games framework, we suggest the user-centered design (UCD) as a design methodology. This iterative methodology is oriented to take into account what the users themselves expect from the game [37]. Finally, compared with commercial games, serious video games must be scientifically tested to check that the usability and efficiency are adequate.

The UCD methodology also allows for recording user performance; collecting this preliminary data with real users allows for closer adjustment to the patient's needs. In the development of TSTM [21,22], there were two events that emerged in user testing. Enigma had to be refined to improve the impact on working memory, and for Smasher, including a bug representing a distractor (a bird flying in a strange way), psychologists suggested that this incoherent stimulus could be more distracting and was incorporated as an advanced distractor. UCD methodology allowed us to redesign the game mechanics considering how people with ADHD behaved during the game.

The initial TSTM design was modified following suggestions from patients with ADHD about esthetics, game difficulty, or rewards [22]. The UCD methodology fulfilled two additional objectives: the designers checked the correct functioning of the game and the ADHD patients felt involved in the development process.

Engaging a Patient With ADHD to Begin Treatment

Treatment with therapeutic video games may help to improve adherence to multimodal treatment in ADHD by bringing patients closer to treatment.

The game's theme is important in the design process, as player preferences may differ from one player to another. Games such as EndeavorRx [38] or Plan-it-Commander [39] are inspired by space and astronauts. Regarding TSTM, each mini game takes place in the forest, as regular interaction with nature can be beneficial for people with ADHD [40]. Furthermore, it is possible to design exercises that are compatible with different game environments. Thus, users could choose the setting of the game without modifying its mechanics. By doing so, motivation for the task may be enhanced along with an increase in the adherence to treatment. Although this guide is oriented to the usually chosen age range of 6 to 18 years, generating different appearances depending on the patient's age would likely be a good approach to treat ADHD during an individual's growth and development. The same mechanics could even be used for adults by adopting more mature and sophisticated themes.

A serious video game also has the following requirements: (1) easy to understand, (2) require a minimal cognitive load (baseline), and (3) gradual level of difficulty. The baseline can be designed as an in-game tutorial. It is important to ensure that the person has understood the task and how to perform it. All of these elements were examined in the previously mentioned usability study for TSTM [22]. The aforementioned user-centered model [37] can help to adjust design guidelines.

Keeping a Patient Motivated With Serious Video Games

Sampayo-Vargas et al [41] stressed that in order to make educational computer games intrinsically motivational, it is critical to provide an optimal level of challenge. The difficulty curve should be neither too boring nor too frustrating. Users play the game and try to learn the correct pattern to follow to pass the given task and master it. Initially, they are entertained; however, when the game is no longer a challenge, users tend to become bored and give up [42]. For people with ADHD, novel challenges may improve adherence to treatment by increasing motivation. Furthermore, the game must automatically adjust to the individual's performance level [37,41]. Games such as EndeavorRx [38] change the level of difficulty until the player is performing at an 80% rate of accuracy. Here, it is important to stress that artificial intelligence (AI) may help to personalize and update the needs of each individual in a specific context and time by adjusting at that precise moment (eg, type of game, intensity).

To avoid boredom, game design components should be included in any serious video game related to objects, mechanics, dynamics, and emotions [43]. "Juiciness" is a novel concept that refers to adding visual embellishments, sounds, and other types of nonfunctional elements, thus allowing developers to improve players' experience in their games [29]. Juiciness should not affect the game mechanics but rather only focus on the extra elements (eg, menus, transitions, feedback), pursuing the development of a simple, minimalistic, intuitive, and easy-to-understand interface.

People with ADHD have limited self-perception and objectivity in task performance [4]. Therefore, providing simple and direct feedback as the player progresses consistently is a fundamental requirement in the design of the proposed serious video games [42]. To avoid too much frustration that may lead to abandoning the video game, points, badges, leaderboards, and other game components can be used. Moreover, introducing small doses of challenge could be helpful to improve frustration tolerance. Achieving engagement through the attractiveness of the task may benefit emotional dysregulation. Indeed, a recent review [44] concluded that regular, nonexcessive video game use can help to enhance emotional regulation in children. A good

XSL•FO
RenderX

example is EmoGalaxy [45], which is a video game that improved emotional regulation in children diagnosed with ADHD.

Although video games are usually visually based, music and sounds are also fundamental elements that facilitate flow in video games alongside other elements. However, these elements should not be overwhelming, because people with ADHD are more likely to be distracted. Thus, audio-visual integration can be used to vary the intensity of emotional experience states [46]. To facilitate the sense of progress, missions can be completed. Performance can be quantified and made visible in changes in the user interface, counters, and score, among other elements [42]. This feedback can be provided by using as many senses as possible, including visual stimulus (ie, colors to show the level of performance), auditory stimulus (ie, sounds to point at a hit, error, task change), or nociceptive stimulus (ie, the vibration of the game controller can indicate that a mistake has been made).

Improving ADHD Symptoms

People with ADHD present broad clinical and cognitive variability [3]. Accordingly, any serious therapeutic intervention should ideally provide a tailored treatment targeting each patient's specific difficulties and circumstances. In other words, serious video games should be developed within a personalized medicine framework [47].

Dovis et al [48] suggested that children with ADHD are less stimulated by reinforcement than typically developing children (likely due to a dopaminergic deficit), and therefore require higher amounts and frequencies of reward to perform optimally. These rewards should be more frequent at the beginning of the game (extrinsic motivation) and should decrease in frequency as the sessions progress, rewarding good performance with larger rewards (intrinsic motivation). However, people with ADHD may have delay aversion: a preference for small, immediate over larger, but delayed rewards [49]. Thus, it may be interesting to work on impulsivity through the reinforcement system [50]. Good task execution can also be rewarded through juicy embellishments and in-game achievements.

Avoiding Addiction

As previously discussed, people with ADHD may be more dependent on extrinsic motivation [51,52]. Children with ADHD are also more at risk of addiction to video games [13]. To avoid this, the focus should be on limiting the time spent playing. Furthermore, these time constraints may positively impact engagement by avoiding fatigue.

Research estimates that the average student's attention span is 10-20 minutes [53]; however, one of the symptoms of ADHD is sustained attention. Children with ADHD frequently experience hyper focus (the experience of focusing on something to the extent that all other stimuli are almost completely excluded) when playing video games [54]. Despite this, hyper focus cannot ensure that patients will continue to receive benefits, as illustrated in theories of deliberate practice [55].

Consequently, we have to reach a compromise between reinforcing the patient's attention beyond their own attention span and reducing gaming time to avoid addiction. Following the state of the art [15-17], treatments of 20-40 minutes 4-5 days a week seem to achieve more positive effects.

King et al [56] proposed five characteristics and their potential for game abuse: social features, manipulation and control features, narrative and identity features, reward and punishment features, and presentation features (see Table 1).

Table 1. Game characteristics and potential for game abuse (based on King et al [56]).

Game feature	Potential for game abuse
Social features	Do not play a relevant role in serious games, because they tend to be individually played
Control features	Need to master game controls
Resource management	Keeping resources above a minimum amount
Reward/punishment	Bonus content or additional resources associated with outstanding performance
Narrative and identity	Excess of immersion, excess of attachment to player's avatar

Sessions should last approximately 20-30 minutes depending on the patient's age. We recommend dividing the session into a variety of tasks [52]. Every 10 minutes, there can be a change of task (greater variability) or short breaks [4]. Furthermore, it is important to plan a sufficient number of sessions to consolidate learning. For instance, each TSTM training session lasts 25-30 minutes divided into three blocks. Two blocks of 10 minutes each are used for training with two game mechanics (Smasher, Kuburi, Tekateki, or Enigma), along with a third block of 5 minutes of chess-based game mechanics. The sessions are counterbalanced, randomizing the order of the blocks to avoid being repetitive or boring (see [23]).

Game Mechanics

According to Koster's [57] definition in the Theory of Fun of Game Design, game mechanics are "rule-based systems/simulations that facilitate and encourage the user to explore and learn the properties of their possibility space through the use of feedback mechanisms."

In a serious video game, game mechanics are the main intervention components through which patients acquire metacognitive strategies and skills to improve their cognitive abilities [18]. Game mechanics promote learning and repetition of the task, which can help to establish generalization.

Analyzing the serious video games included in several recent reviews [15-17] led us to the conclusion that almost all current ADHD treatments are structured in a compilation of mini games

XSL•FO
RenderX

that generally tend to be similar to a regular video game (with respect to narrative, lore, esthetics, etc). In contrast, serious games for ADHD assessment are rarely more than one mini game, inspired by single tests. This is in keeping with a general recommendation of developing short and discrete tasks for people with ADHD, among others [43]. The use of mini games provides a clear advantage: the combination of mini games helps to personalize treatment for each patient by being able to distinguish which cognitive processes require more training.

Mini games can be created by adapting a cognitive task using game-up and mapping techniques [43] or from scratch. To attract attention, we can introduce design strategies such as gamifying the cognitive task stimuli to try to produce positive sensations. However, this process is specific to the disorder for which the treatment is intended. Khaleghi et al [43] proposed following the Objects, Mechanics, Dynamics, Emotions (OMDE) design guidelines. When designing game mechanics, it is critical to validate the cognitive aspects of the game. For example, in people with ADHD, it is necessary to control distractions, the cognitive load, and the time spent. Indeed, there are three main cognitive aspects that affect people with ADHD: attention, executive functions, and daily skills.

Inattention is one of the core symptoms of ADHD [2]. In traditional attention-training tasks, children with ADHD perform worse than typically developing children [52]. New mini games can be created to treat different types of attention. Additionally, ADHD presents great comorbidity with executive dysfunction [3]. Some games can help to train executive functions such as inhibition control, working memory, planning, or reasoning.

Lastly, one of the major criticisms of the use of serious video games is the lack of ecological validity [39,40]. Thus, proposed video games should induce the transfer of cognitive skills to domains of daily life functioning (transferability). Transferability is based in the improvement of different executive functions, which may help with scholarly development and motivation. Cognitive training is effective if the patient applies it in their daily life. For example, children with ADHD show impairments in social skills, which can have serious long-term consequences. Games that explore cooperation skills, such as “Space Club” [39], may improve social benefits. For example, TSTM incorporated chess mechanics, because chess can improve math performance, thus demonstrating transference into the educational domain [21].

Exploration

The exploration stage of the IDEAL-Games framework [19] ensures that the video game meets the therapeutic objectives by testing the game’s feasibility, usability, reliability, and validity using proof-of-concept and pilot studies [39,43]. If not satisfactory, the development team can go back to the design phase, and revise the objectives, theoretical model, target variables, and players’ experience [18]. This stage should verify that the serious video game is ready (or not) for testing effectiveness. However, before testing the effectiveness, it is critical to verify that users enjoy the game [18]. A video game could be effective for treating ADHD, but will not be useful if patients do not use it because it is boring. Thus, feedback from patients may help to redesign certain elements. Furthermore, proof of concept, usability, and validity research studies help to detect bugs that may interfere with the following stages and to identify design errors, which would mean going back to the design stage and avoid conducting clinical studies that are destined to fail.

In the video game TSTM, we used UCD methodology [22,37] to follow precision gaming features, difficulty parameters, and other elements highlighted above. Feasibility was studied by conducting a market survey with professionals. The usability study of TSTM was assessed in people with ADHD to verify that the game was usable, of interest to the target group, easy to understand, and without side effects [22]. The reliability analysis allowed exploring the permanence of the measure over time (test-retest) or between raters (interrater), the consistency of the game (eg, with difficulty levels), or to determine whether two versions of the game (with and without VR) have similar results (parallel-forms reliability). In terms of validity, TSTM was considered to have content validity from the very beginning. For ecological validity, we monitored the transferability.

Assessment of Effectiveness

Assessment is critical to determine the effectiveness of a serious video game [40]. Although research on health-based games has increased, the scientific basis remains too poor for a generalized recommendation [19]. High-quality studies (ie, randomized clinical trials) with sufficient sample size and time for the intervention are still warranted [43]. The key methodological limitations frequently encountered are inadequate controls, sustainability of training, and measures of cognitive function or inappropriate behavior [40]. Accordingly, we conducted a randomized trial with TSTM [23,58]. The patients’ self-reports demonstrated significant improvements in emotional regulation and adaptation to the social context (positive impression and interpersonal scale); 96% of the patients reported liking the experience of the clinical trial with the video game and having perceived a sense of improvement in all the cognitive abilities worked on (especially working memory and impulse control). The face validity also seemed to show that this treatment is more motivating owing to lower dropouts [58].

Long-Term Assessment

As mentioned above, studies sometimes present difficulties in demonstrating transfer and ecological validity. Although some studies demonstrated that playing video games for 16 hours already produce certain changes in the brain [59], it is essential to understand the aspects of the game that are necessary for this to occur, and for how long these positive effects last. Monitoring the long-term effect of brand-new treatments such as video games is imperative, even if their short-term benefit has been clearly established.

Discussion

Summary

This guide aims to provide a roadmap for video game developers and researchers to build a serious video game that is fun, motivating, and effective in treating ADHD, while avoiding addictive properties. Following the IDEAL-Games proposal [19] and suggestions from Duncan et al [18] and Baranowski

XSL•FO
RenderX

[20], among others, the proposed guide tries to bring together current knowledge on this topic. We have followed the order of the IDEAL process to develop the most important sections of this guide. We used our own experience in developing and testing TSTM, a serious video game specifically designed to treat ADHD. Both this guide and TSTM are based on information derived from game design methodologies, contributions from clinicians' expertise, designers' experience, children's interests, and family feedback. TSTM usability results are good [22].

Strengths and Limitations

The use of video games for treating ADHD has many advantages: (1) it is cheap, comfortable, flexible, and cost-efficient (eg, the time saved by avoiding travel alone would also indirectly benefit the environment); (2) the use of novel platforms (ie, VR/alternate reality) may improve not only adherence but also the clinical effects if transferability is demonstrated; and (3) the incorporation of AI may offer more personalized treatments for each patient.

Although serious video games promise to be an effective and interesting treatment platform, several limitations should be mentioned. For instance, serious video games must be tailored to the singularity of every patient (personalized medicine). We must also consider the possibility of some conditions such as motor difficulties, visual impairments, dyslexia, or color blindness, among others that may compromise effectiveness, and introduce adaptations if necessary. In addition, patients may be unfamiliar with technological platforms. Accordingly, a gradual approach is recommended. Furthermore, technological platforms could have side effects [22]. For example, VR can cause dizziness or discomfort in some users. Indeed, VR technology is not recommended in people with epilepsy or children aged 12 years or below. Another limitation is addiction [13]; thus, monitoring this risk is important [51]. Finally, follow-up studies are mandatory before serious video games can be incorporated within the multimodal treatment of ADHD.

Conclusion and Future Directions

We hope that this guide assists the development of serious video games for ADHD. We stress the relevance of having a multidisciplinary team to balance clinical and technical skills [31]. Furthermore, finding a balance between enjoyment and addiction is mandatory. Moreover, new AI techniques may assist in creating tailored cognitive training schedules for each single patient. A virtual tutor may also help children to adhere to treatment by providing positive reinforcement while being monitored by therapists in parallel.

Despite the lack of studies comparing the effectiveness of commercial and serious video games on ADHD, we consider that serious video games have several advantages such as the specific design for symptoms and defective executive functions, although preventing the development of an addiction to the video game is essential.

Finally, evidence-based video serious games could help to complement multimodal treatment. Currently, the diagnosis of ADHD is based on clinical criteria, objective tests, and the subjective view of caregivers and teachers. However, recent studies suggest that video games may also assist in improving ADHD diagnoses. Future actions such as incorporating new technologies (ie, eye tracking, body trackers, brain-computer interfaces) while using a serious video game may help to improve the current subjective diagnoses of ADHD.

Acknowledgments

This research was partially funded by the Spanish National Project (grant number RTI2018-101857-B-I00), Fundación para la Innovación y la Prospectiva en Salud en España to HBF, Centro para el Desarrollo Tecnológico Industrial (European Regional Development Fund IDI-20180701, file 00107278) to MMM, Instituto de Salud Carlos III (AES2021, DTS21/00091) to DDG, Contrato Predoctoral de Formación en Investigación en Salud (i-PFIS contract IFI16/00039) to MRY, Universidad Carlos III de Madrid grant for the requalification of permanent lectures to DDG, and Comunidad de Madrid (PEJD-2019-PRE/SAL-17036 to AS and IND2020/BMD-17544 to MMM).

Authors' Contributions

Conceptualization, AS and MMM; software, AS, MMM, CGT, and MRY; investigation, AS, MMM, and MBF; writing—original draft preparation, AS, MMM, and MBF; writing—review and editing, AS, MMM, MBF, CGT, MRY, DDG, and HBF; supervision, HBF; project administration, HBF; funding acquisition, MMM and HBF. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

In the last 24 months, HBF received lecture fees from Shire. He is a recipient of an IDIPHIPSA intensification grant; involved in two clinical trials (MENSIA KOALA, NEWROFEED Study; ESKETSUI2002); and a member of the Advisory Board of ITA Salud Mental. The remaining authors do not have any conflict of interest regarding the publication of this manuscript.

References

1. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry 2018 Feb;5(2):175-186. [doi: 10.1016/S2215-0366(17)30167-0] [Medline: 29033005]

XSL•FO
RenderX

2. American Psychaitric Association. Diagnostic and Statistical Manual of mental disorders (5th Edition). Washington, DC: American Psychiatric Association; 2013.
3. Castellanos FX, Sonuga-Barke EJ, Milham MP, Tannock R. Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cogn Sci 2006 Mar;10(3):117-123. [doi: 10.1016/j.tics.2006.01.011] [Medline: 16460990]
4. Barkley R. Attention-deficit hyperactivity disorder: a handbook for diagnosis and treatment. Fourth Edition. New York, NY: Guilford Press; 2018.
5. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet 2015 May 30;385(9983):2190-2196. [doi: 10.1016/S0140-6736(14)61684-6] [Medline: 25726514]
6. Dalrymple RA, McKenna Maxwell L, Russell S, Duthie J. NICE guideline review: Attention deficit hyperactivity disorder: diagnosis and management (NG87). Arch Dis Child Educ Pract Ed 2020 Oct 27;105(5):289-293. [doi: 10.1136/archdischild-2019-316928] [Medline: 31776172]
7. Marcus SC, Durkin M. Stimulant adherence and academic performance in urban youth with attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2011 May;50(5):480-489. [doi: 10.1016/j.jaac.2011.02.007] [Medline: 21515197]
8. Toomey SL, Sox CM, Rusinak D, Finkelstein JA. Why do children with ADHD discontinue their medication? Clin Pediatr (Phila) 2012 Aug 14;51(8):763-769. [doi: 10.1177/0009922812446744] [Medline: 22584541]
9. Antshel KM, Faraone SV, Gordon M. Cognitive behavioral treatment outcomes in adolescent ADHD. J Atten Disord 2014 Aug;18(6):483-495. [doi: 10.1177/1087054712443155] [Medline: 22628140]
10. Pottegård A, Bjerregaard BK, Kortegaard LS, Zoëga H. Early discontinuation of attention-deficit/hyperactivity disorder drug treatment: a Danish nationwide drug utilization study. Basic Clin Pharmacol Toxicol 2015 Apr 07;116(4):349-353. [doi: 10.1111/bcpt.12325] [Medline: 25220739]
11. Sonuga-Barke EJ. The dual pathway model of AD/HD: an elaboration of neuro-developmental characteristics. Neurosci Biobehav Rev 2003 Nov;27(7):593-604. [doi: 10.1016/j.neubiorev.2003.08.005] [Medline: 14624804]
12. Data and key facts. Interactive Software Federation for Europe (ISFE). URL: https://www.isfe.eu/data-key-facts/ [accessed 2022-04-04]
13. Menéndez-García A, Jiménez-Arroyo A, Rodrigo-Yanguas M, Marín-Vila M, Sánchez-Sánchez F, Román-Riechmann E, et al. Internet, video game and mobile phone addiction in children and adolescents diagnosed with ADHD: a case-control study. Adicciones 2020 Dec 04:1469. [doi: 10.20882/adicciones.1469] [Medline: 33338245]
14. Rapport MD, Orban SA, Kofler MJ, Friedman LM. Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. Clin Psychol Rev 2013 Dec;33(8):1237-1252. [doi: 10.1016/j.cpr.2013.08.005] [Medline: 24120258]
15. Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, Delgado-Gomez D, Navarro-Jimenez R, Baca-Garcia E, et al. Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. Eur Child Adolesc Psychiatry 2022 Jan 18;31(1):5-20. [doi: 10.1007/s00787-020-01557-w] [Medline: 32424511]
16. Zheng Y, Li R, Li S, Zhang Y, Yang S, Ning H. A review on serious games for ADHD. arXiv. 2021 Mar. URL: https://arxiv.org/pdf/2105.02970.pdf [accessed 2022-04-04]
17. Rodrigo-Yanguas M, González-Tardón C, Bella-Fernández M, Blasco-Fontecilla H. Serious video games: angels or demons in patients with attention-deficit hyperactivity disorder? A quasi-systematic review. Front Psychiatry 2022;13:798480. [doi: 10.3389/fpsyt.2022.798480] [Medline: 35573357]
18. Duncan LR, Hieftje KD, Culyba S, Fiellin LE. Game playbooks: tools to guide multidisciplinary teams in developing videogame-based behavior change interventions. Transl Behav Med 2014 Mar;4(1):108-116 [FREE Full text] [doi: 10.1007/s13142-013-0246-8] [Medline: 24653781]
19. Kinross JM. Precision gaming for health: Computer games as digital medicine. Methods 2018 Dec 01;151:28-33. [doi: 10.1016/j.ymeth.2018.09.009] [Medline: 30273711]
20. Baranowski T. Descriptions for articles introducing a new game for health. Games Health J 2014 Apr;3(2):55-56. [doi: 10.1089/g4h.2014.0008] [Medline: 26196042]
21. Rodrigo-Yanguas M, Martín-Moratinos M, González Tardón C, Blasco-Fontecilla H. Virtual reality and chess. A video game for cognitive training in patients with ADHD. 2020 Presented at: VI Congreso de la Sociedad Española para las Ciencias del Videojuego; October 2020; Online p. 7-8 URL: http://ceur-ws.org/Vol-2719/paper20.pdf
22. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Royuela A, Blasco-Fontecilla H. A virtual reality game (The Secret Trail of Moon) for treating attention-deficit/hyperactivity disorder: development and usability study. JMIR Serious Games 2021 Sep 01;9(3):e26824 [FREE Full text] [doi: 10.2196/26824] [Medline: 34468332]
23. Rodrigo-Yanguas M, Martin-Moratinos M, Menendez-Garcia A, Gonzalez-Tardon C, Sanchez-Sanchez F, Royuela A, et al. A virtual reality serious videogame versus online chess augmentation in patients with attention deficit hyperactivity disorder: a randomized clinical trial. Games Health J 2021 Aug;10(4):283-292. [doi: 10.1089/g4h.2021.0073] [Medline: 34370610]
24. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. Acta Psychol 2008 Nov;129(3):387-398. [doi: 10.1016/j.actpsy.2008.09.005] [Medline: 18929349]

XSL•FO
RenderX

25. Granic I, Lobel A, Engels RCME. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
26. Nuyens FM, Kuss DJ, Lopez-Fernandez O, Griffiths MD. The empirical analysis of non-problematic video gaming and cognitive skills: a systematic review. Int J Ment Health Addiction 2018 Jun 20;17(2):389-414. [doi: 10.1007/s11469-018-9946-0]
27. Pine R, Fleming T, McCallum S, Sutcliffe K. The effects of casual videogames on anxiety, depression, stress, and low mood: a systematic review. Games Health J 2020 Aug;9(4):255-264. [doi: 10.1089/g4h.2019.0132] [Medline: 32053021]
28. Michael D, Chen S. Serious games: Games that educate, train, and inform. Cincinnati, OH: Muska & Lipman/Premier-Trade; 2005.
29. Swink S. Game feel: a game designer's guide to virtual sensation. Boca Raton, FL: CRC Press; 2008.
30. Evans MA, Norton A, Chang M, Deater-Deckard K, Balci O. Youth and video games. Zeitschrift Psychologie 2013 Jan;221(2):98-106. [doi: 10.1027/2151-2604/a000135]
31. Malinverni L, Mora-Guiard J, Padillo V, Valero L, Hervás A, Pares N. An inclusive design approach for developing video games for children with autism spectrum disorder. Comput Hum Behav 2017 Jun;71:535-549. [doi: 10.1016/j.chb.2016.01.018]
32. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [doi: 10.1037/0033-2909.121.1.65] [Medline: 9000892]
33. Brown TE. ADD/ADHD and impaired executive function in clinical practice. Curr Atten Disord Rep 2009 Mar 27;1(1):37-41. [doi: 10.1007/s12618-009-0006-3] [Medline: 18803914]
34. Moreau D, Conway AR. The case for an ecological approach to cognitive training. Trends Cogn Sci 2014 Jul;18(7):334-336. [doi: 10.1016/j.tics.2014.03.009] [Medline: 24972505]
35. Benzing V, Schmidt M. The effect of exergaming on executive functions in children with ADHD: a randomized clinical trial. Scand J Med Sci Sports 2019 Aug 23;29(8):1243-1253. [doi: 10.1111/sms.13446] [Medline: 31050851]
36. Lim CG, Lee TS, Guan C, Fung DSS, Zhao Y, Teng SSW, et al. A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder. PLoS One 2012 Oct 24;7(10):e46692 [FREE Full text] [doi: 10.1371/journal.pone.0046692] [Medline: 23115630]
37. Fullerton T. Game design workshop: a playcentric approach to creating innovative games. Burlington, MA: Morgan Kaufmann Publishers; 2018.
38. Davis NO, Bower J, Kollins SH. Proof-of-concept study of an at-home, engaging, digital intervention for pediatric ADHD. PLoS One 2018 Jan 11;13(1):e0189749 [FREE Full text] [doi: 10.1371/journal.pone.0189749] [Medline: 29324745]
39. Bul KCM, Franken IHA, Van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, et al. Development and user satisfaction of "Plan-It Commander," a serious game for children with ADHD. Games Health J 2015 Dec;4(6):502-512. [doi: 10.1089/g4h.2015.0021] [Medline: 26325247]
40. Rabipour S, Raz A. Training the brain: fact and fad in cognitive and behavioral remediation. Brain Cogn 2012 Jul;79(2):159-179. [doi: 10.1016/j.bandc.2012.02.006] [Medline: 22463872]
41. Sampayo-Vargas S, Cope CJ, He Z, Byrne GJ. The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. Comput Educ 2013 Nov;69:452-462. [doi: 10.1016/j.compedu.2013.07.004]
42. Murphy C, Chertoff D, Guerrero M, Moffitt K. Design better games: flow, motivation, and fun in learning games. In: Hussain TS, Coleman SL, editors. Design and development of training games: practical guidelines from a multidisciplinary perspective. Cambridge, UK: Cambridge University Press; 2015:146-178.
43. Khaleghi A, Aghaei Z, Mahdavi MA. A gamification framework for cognitive assessment and cognitive training: qualitative study. JMIR Serious Games 2021 May 18;9(2):e21900 [FREE Full text] [doi: 10.2196/21900] [Medline: 33819164]
44. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: a systematic review. Games Health J 2018 Apr;7(2):85-99. [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
45. Hakimirad E, Kashani-Vahid L, Hosseini MS, Irani A, Moradi H. Effectiveness of EmoGalaxy video game on social skills of children with ADHD. 2019 Presented at: 2019 International Serious Games Symposium (ISGS); December 26, 2019; Tehran, Iran p. 7-12. [doi: 10.1109/isgs49501.2019.9046992]
46. Pan F, Zhang L, Ou Y, Zhang X. The audio-visual integration effect on music emotion: behavioral and physiological evidence. PLoS One 2019 May 30;14(5):e0217040 [FREE Full text] [doi: 10.1371/journal.pone.0217040] [Medline: 31145745]
47. de Leon J. Focusing on drug versus disease mechanisms and on clinical subgrouping to advance personalised medicine in psychiatry. Acta Neuropsychiatr 2014 Dec 19;26(6):327-333. [doi: 10.1017/neu.2014.14] [Medline: 25455256]
48. Dovis S, Van der Oord S, Wiers RW, Prins PJM. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
49. Sjöwall D, Roth L, Lindqvist S, Thorell LB. Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. J Child Psychol Psychiatry 2013 Jun 15;54(6):619-627. [doi: 10.1111/jcpp.12006] [Medline: 23061803]
50. Engelstein G. Achievement relocked: Loss aversion and game design. Cambridge, MA: MIT Press; 2020.

XSL•FO
RenderX

51. Mathews CL, Morrell HER, Molle JE. Video game addiction, ADHD symptomatology, and video game reinforcement. Am J Drug Alcohol Abuse 2019 Jun 06;45(1):67-76. [doi: 10.1080/00952990.2018.1472269] [Medline: 29874473]
52. Skalski S, Pochwatko G, Balas R. Impact of motivation on selected aspects of attention in children with ADHD. Child Psychiatry Hum Dev 2021 Aug 20;52(4):586-595 [FREE Full text] [doi: 10.1007/s10578-020-01042-0] [Medline: 32816140]
53. Tokuhama-Espinosa T. Mind, brain, and education science: a comprehensive guide to the new brain-based teaching. New York, NY: W. W. Norton & Company; 2010.
54. Ashinoff BK, Abu-Akel A. Hyperfocus: the forgotten frontier of attention. Psychol Res 2021 Feb 20;85(1):1-19 [FREE Full text] [doi: 10.1007/s00426-019-01245-8] [Medline: 31541305]
55. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev 1993;100(3):363-406. [doi: 10.1037/0033-295x.100.3.363]
56. King D, Delfabbro P, Griffiths M. Video game structural characteristics: a new psychological taxonomy. Int J Ment Health Addiction 2009 Apr 7;8(1):90-106. [doi: 10.1007/s11469-009-9206-4]
57. Koster R. Theory of fun for game design. Sebastopol, CA: O'Reilly Media; 2013.
58. Rodrigo-Yanguas M. Effectiveness of a personalized, chess-based training serious video game in the treatment of children and adolescents with attention deficit/hyperactivity disorder (ADHD): a randomized clinical study. Games For Health Journal 2022:in review.
59. Brilliant TD, Nouchi R, Kawashima R. Does video gaming have impacts on the brain: evidence from a systematic review. Brain Sci 2019 Sep 25;9(10):251 [FREE Full text] [doi: 10.3390/brainsci9100251] [Medline: 31557907]

Abbreviations

ADHD: attention deficit hyperactivity disorder
AI: artificial intelligence
CBT: cognitive behavioral therapy
OMDE: Objects, Mechanics, Dynamics, Emotions
TSTM: The Secret Trail of Moon
UCD: user-centered design
VR: virtual reality

Edited by N Zary; submitted 28.09.21; peer-reviewed by A Khaleghi, A Bikic, D Frohlich; comments to author 17.02.22; revised version received 12.04.22; accepted 12.06.22; published 01.08.22.

Please cite as:
Súja r A, Martín-Moratinos M, Rodrigo-Yanguas M, Bella-Fernández M, González-Tardón C, Delgado-Gómez D, Blasco-Fontecilla H
Developing Serious Video Games to Treat Attention Deficit Hyperactivity Disorder: Tutorial Guide
JMIR Serious Games 2022;10(3):e33884
URL: https://games.jmir.org/2022/3/e33884
doi:10.2196/33884
PMID:35916694

©Aarón Súja r, Marina Martín-Moratinos, María Rodrigo-Yanguas, Marcos Bella-Fernández, Carlos González-Tardón, David Delgado-Gómez, Hilario Blasco-Fontecilla. Originally published in JMIR Serious Games (https://games.jmir.org), 01.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Social Media Users' Perceptions of a Wearable Mixed Reality Headset During the COVID-19 Pandemic: Aspect-Based Sentiment Analysis

Heejin Jeong[1,2], PhD; Allison Bayro[2], BS; Sai Patipati Umesh[3]; Kaushal Mamgain[3], MS; Moontae Lee[4], PhD

[1]Department of Mechanical and Industrial Engineering, University of Illinois Chicago, Chicago, IL, United States

[2]Richard and Loan Hill Department of Biomedical Engineering, University of Illinois Chicago, Chicago, IL, United States

[3]Department of Computer Science, University of Illinois Chicago, Chicago, IL, United States

[4]Department of Information and Decision Sciences, University of Illinois Chicago, Chicago, IL, United States

Corresponding Author:
Heejin Jeong, PhD
Department of Mechanical and Industrial Engineering
University of Illinois Chicago
842 W Taylor St
Chicago, IL, 60607
United States
Phone: 1 312 355 5558
Email: heejinj@uic.edu

Abstract

Background: Mixed reality (MR) devices provide real-time environments for physical-digital interactions across many domains. Owing to the unprecedented COVID-19 pandemic, MR technologies have supported many new use cases in the health care industry, enabling social distancing practices to minimize the risk of contact and transmission. Despite their novelty and increasing popularity, public evaluations are sparse and often rely on social interactions among users, developers, researchers, and potential buyers.

Objective: The purpose of this study is to use aspect-based sentiment analysis to explore changes in sentiment during the onset of the COVID-19 pandemic as new use cases emerged in the health care industry; to characterize net insights for MR developers, researchers, and users; and to analyze the features of HoloLens 2 (Microsoft Corporation) that are helpful for certain fields and purposes.

Methods: To investigate the user sentiment, we collected 8492 tweets on a wearable MR headset, HoloLens 2, during the initial 10 months since its release in late 2019, coinciding with the onset of the pandemic. Human annotators rated the individual tweets as positive, negative, neutral, or inconclusive. Furthermore, by hiring an interannotator to ensure agreements between the annotators, we used various word vector representations to measure the impact of specific words on sentiment ratings. Following the sentiment classification for each tweet, we trained a model for sentiment analysis via supervised learning.

Results: The results of our sentiment analysis showed that the bag-of-words tokenizing method using a random forest supervised learning approach produced the highest accuracy of the test set at 81.29%. Furthermore, the results showed an apparent change in sentiment during the COVID-19 pandemic period. During the onset of the pandemic, consumer goods were severely affected, which aligns with a drop in both positive and negative sentiment. Following this, there is a sudden spike in positive sentiment, hypothesized to be caused by the new use cases of the device in health care education and training. This pandemic also aligns with drastic changes in the increased number of practical insights for MR developers, researchers, and users and positive net sentiments toward the HoloLens 2 characteristics.

Conclusions: Our approach suggests a simple yet effective way to survey public opinion about new hardware devices quickly. The findings of this study contribute to a holistic understanding of public perception and acceptance of MR technologies during the COVID-19 pandemic and highlight several new implementations of HoloLens 2 in health care. We hope that these findings will inspire new use cases and technological features.

(JMIR Serious Games 2022;10(3):e36850) doi:10.2196/36850

XSL•FO
RenderX

KEYWORDS

HoloLens 2; sentiment analysis; natural language processing, Twitter; COVID-19; usability evaluation

Introduction

Background

The release of new virtual reality (VR), augmented reality (AR), or mixed reality (MR) devices elicits a global conversation between VR, AR, and MR developers and users through social media. Such public views may significantly influence the future purchases of potential customers including users, developers, and researchers. Thus, it is essential and meaningful to investigate these views about their usage. This was especially crucial during the unprecedented COVID-19 pandemic, when MR technologies enabled socially distanced education and training in the health care industry. Furthermore, such viewpoints inspire new use cases, which influence health care policy interventions. This investigation offers insights into potential application areas, strengths and weaknesses, and product improvements for future releases. These insights derived from consumer perceptions serve as feedback for the curators to experiment and enhance product capabilities and expand on new use cases inspired by the pandemic.

Previous studies have evaluated the usability and sentiment of VR, AR, and MR headsets [1-3], but there are some limitations. First, there is a lack of evaluations that analyze the usability of sentiments for developers, researchers, and users separately [4]. Moreover, most studies have been evaluated with a limited number of people invited to the laboratory [2,5,6]. Finally, the real-time opinions worldwide have not been reflected [4]. In this study, we propose aspect-based sentiment analysis using Twitter-derived tweets to complement the shortcomings of the existing usability evaluations.

The focus of this study was to explore the usability and sentiment of 1 representative MR headset, Microsoft HoloLens 2, launched in November 2019. HoloLens 2 is the successor product of the initial version released in March 2016. A summary of the comparison between the 2 versions of the HoloLens devices is shown in Table 1. HoloLens 2 has some significant developments compared with the first model. These added developments and features contribute to overall user sentiment. It has new eye-tracking features and gestures. Furthermore, it also has better depth detection, better memory storage, a modern Bluetooth connection, an improved USB port, and a more powerful RAM. Eye tracking enables developers to measure the point of gaze, which benefits eye gaze–based interactions. Kościesza [7] reported that the gesture sensors can recognize up to 25 points of articulation from the fingers and wrist enabling refined object manipulation. In addition, HoloLens 2 also offers a better resolution and field of view. This allows the users to see more without having to turn their heads. Ergonomically, the device also has a knob to enable resizing capabilities for the best fit. A small change in weight makes it slightly more comfortable to wear for a longer duration. The visor flips up, allowing users to wear glasses inside if needed. Thus, HoloLens 2 specifications enable users to manipulate holograms easily and can be used by people of all skill levels for various applications.

Table 1. Comparison of HoloLens 1 and 2 (adapted from Kościesza [7] and recreated).

Specification	HoloLens 1	HoloLens 2
Display resolution	1280×720 pixels (per eye)	2048×1080 pixels (per eye)
Field of view	34°	52°
Weight	579 g	566 g
Camera	2.4 MP, HD video	8 MP stills, 1080p video
Audio	Built-in speakers; 3.5-mm jack	Built-in spatial sound; 3.5-mm jack
Built-in microphone	4-microphone array	5-microphone array
Voice command	Yes	Yes
Eye tracking	No	Yes
Biometric security	No	Yes
Hand tracking	1 hand	2 hands full tracking
Price	US $3000	US $3500 or US $99-125 per month
Gestures: press, grab, direct manipulation, touch interaction, scroll with a wave	No	Yes
Memory and storage	1 GB, 64 GB	4 GB, 64 GB

In this study, we analyzed tweets extracted from November 2019 to August 2020, for the first 10 months after the release of HoloLens 2, coinciding with the onset of the pandemic. The opinions about HoloLens 2 shared on Twitter were classified based on (1) positive or negative indicators that evaluate the usability and sensibility of the MR headset (ie, usability, field of view, motion sickness, comfort, immersion, cost, and development) and (2) whether it is an opinion that gives insight to MR developers, researchers, and users (yes or no).

XSL•FO
RenderX

This study has 4 main contributions. First, through aspect-based sentiment analysis, it was possible to denote which feature of HoloLens 2 is helpful for certain fields and purposes. Second, the proposed usability evaluation may be used to develop new VR, AR, and MR devices. Third, it enables rapid analyses using real-time data extracted worldwide. Finally, it facilitates an analysis of sentiment changes over time, as the use cases of the HoloLens2, especially in health care, expanded with the pandemic.

Previous Work

Usability Evaluation Cases of VR, AR, and MR Devices

VR, AR, and MR devices have gained popularity, and therefore, there is much research regarding the use cases of such devices [4,8]. VR is a fully immersive technology that shuts out the real world and transposes users to a web- or internet-based space [9]. In contrast, AR is defined as a real-time view of the physical world enhanced by adding virtual computer–generated information [10]. Finally, MR blends the physical world features of AR and virtual world features of VR to produce an environment in which real and digital objects coexist and interact [9]. Egliston and Carter [11] investigated the relatability of Oculus, a VR product by Facebook, to the lives and values of individuals. Specifically, the researchers used YouTube comments posted on promotional videos for the Oculus. Yildirim et al [12] compared three different gaming platforms to evaluate the effect of VR on the video game user experience: (1) desktop computer, (2) Oculus Rift, and (3) HTC Vive. The applications of such devices are not limited to the gaming field. For example, Bayro et al [13] evaluated the use of VR head-mounted display-based and computer-based remote collaboration solutions. Wei et al [14] assessed the suitability of Google Glass in surgical settings. A substantial amount of the literature gathered between January 2013 and May 2017 suggested a moderate to high acceptability of incorporating Google Glass within various surgical environments. It is also essential to evaluate the customer base of VR, AR, and MR products to understand the real-world applications of such devices. Rauschnabel et al [15] aimed to see what users' personality traits enable increased willingness to adopt VR technology. The researchers found that consumers who are notably open and emotionally stable are more aware of Google Glass. Furthermore, consumers who recognize the high functional benefits and social conformity of wearables, such as Google Glass, increase technology adoption. A recent study by Ghasemi and Jeong [16] introduced model-based and large-scale video-based remote evaluation tools that could be used to assess the usability of multimodal interaction modalities in MR.

Usability Evaluation Cases of HoloLens 1 and 2

Since the launch of HoloLens 1 and HoloLens 2, research has suggested some good use cases across domains. Hammady et al [17] studied how HoloLens provides a good experience when used in museums. This study highlighted the restricted field of view in HoloLens and offered an innovative methodology to improve the accessibility of the spatial UI system, thus resulting in a positive user experience. Hoover et al [18] evaluated the effects of different hardware for providing instructions during complex assembly tasks. The researchers noted that HoloLens users usually have lower error rates than non-AR users [18]. Xue et al [19] investigated user satisfaction in terms of both interaction and enjoyment with the HoloLens device. A total of 142 participants from 3 industrial sectors, including aeronautics, medicine, and astronautics. The researchers concluded that general computer knowledge positively affects user satisfaction despite unfamiliarity with the HoloLens smart glasses. Bräuer and Mazarakis [20] tested the use of HoloLens to increase motivation in AR order-picking tasks through gamification. The researchers found that the participants found the AR application intuitive and satisfying. Levy et al [21] discovered that HoloLens 2 is more efficient than HoloLens 1. Park et al [22] stated that using HoloLens 2 resulted in reduced variability and elevated the performance of all operators performing CT-guided interventions, positively affecting this sector of the health care industry. Furthermore, Thees et al [23] explored the impact of HoloLens 1 on fostering learning and reducing extraneous cognitive processing. This study showed a significantly lower extraneous cognitive load during a physics laboratory experiment using the HoloLens 1.

Cases of Sentiment Analysis Based on Social Media

Recently, many studies have used Twitter data to perform sentiment analyses [24]. Carvalho and Plastino [25] highlighted the challenge of this analysis because of the short and informal nature of tweets. Guo et al [26] proposed a Twitter sentiment score model, which exhibits a strong prediction accuracy and reduces the computational burden without the knowledge of historical data. The results of this study provided an efficient model of financial market prediction with an accuracy of 97.87%. Chamlertwat et al [27] proposed a microblog sentiment analysis system that automatically analyzes customer opinions derived from the Twitter microblog service. In the past decade, the Internet of Things (IoT) has also gained popularity. Bian et al [28] mined Twitter to evaluate the public opinion of IoT. Specifically, the researchers collected perceptions of the IoT from multiple Twitter data sources and validated these perceptions against Google Trends. Following this, sentiment analysis was performed to gain insights into public opinion toward the IoT. Mittal and Goel [29] examined the causal relationship between public and market sentiments using a large scale of tweets and a stock market index, the Dow Jones values, from June 2009 to December 2009. Venugopalan and Gupta [30] explored tweet-specific features using domain-independent and domain-specific lexicons to analyze consumer sentiment. In addition, Troisi et al [31] performed a sentiment analysis using data from several social media platforms, including Twitter, to evaluate factors that influence university choice. The researchers noted that the main variable motivating such decision was the training offered, followed closely by physical structure, work opportunities, prestige, and affordability. Nanath and Joy [32] explored the factors that affect COVID-19 pandemic–related content sharing on Twitter by performing natural language processing techniques such as emotion and sentiment analyses. The findings showed that tweets with named entities, expression of negative emotions, referenced mental health, optimistic content, and longer length were more likely shared. Nguyen et al [33] evaluated the association between publicly expressed sentiment toward minorities and resulting

XSL•FO
RenderX

birth outcomes. Using Twitter's streaming application programming interface, the collected and analyzed tweets showed that mothers living in states with the lowest positive sentiment toward minorities had the highest prevalence of low birth weights. Gaspar et al [34] used sentiment analysis techniques to examine affective expressions toward the food contamination caused by enterohemorrhagic *Escherichia coli* in Germany in 2011. The findings highlighted diverse attitudes (positive and negative) and perceived outlooks (threat or challenge), thus emphasizing the ability of sentiment analyses to function as a technique for human-based assessment of stressful events.

Although many studies use data sets of several hundred thousand to millions for sentiment analysis, other researchers report significant findings using <10,000 data points. Myslin et al [35] collected 7362 tobacco-related tweets to develop content and sentiment analysis toward tobacco. The findings suggest that the sentiment toward tobacco was more positive than negative, likely resulting from social image, personal experience, and popular tobacco products. Furthermore, Greaves et al [36] used sentiment analysis techniques to categorize 6412 web-based hospital posts as a positive or negative evaluation of their health care. Using machine learning, the researchers observed moderate associations between predictions on whether patients would recommend a hospital and their responses. More recently, Berkovic et al [37] analyzed 149 arthritis-related tweets to identify topics important to individuals with arthritis during the pandemic and explore the sentiment of such tweets. The results revealed several emerging themes including health care experiences, personal stories, links to relevant blogs, discussion of symptoms, advice sharing, positive messages, and stay-at-home messaging. In addition, the sentiment analysis should address negative concerns about medication shortages, symptom burdens, and the desire for reliable information.

There have also been several sentiment analysis studies in the AR and VR domains. For example, Shahzad et al [38] studied user feedback to evaluate the perception of Fitbit Alta HR (Fitbit). The researchers found that most users spoke highly about such a device. El-Gayar et al [39] used social media analysis techniques to analyze and categorize tweets related to major manufacturers of consumer wearable devices. The analysis provided insight into user priorities related to device characteristics, integration, and wearability issues.

Benefits of Wearable MR Technologies in Health Care

With the rapid onset of the COVID-19 pandemic, MR technologies have become a revolutionary tool in the health care industry to support educational endeavors, patient care, and rehabilitation. Martin et al [40] explored the capabilities of MR technology to enable telemedicine to support patient care during the pandemic. This study found that the HoloLens2 facilitated a 51.5% reduction in health care workers (HCWs) time exposure to patients with COVID-19 and an 83.1% reduction in the amount of personal protective equipment (PPE) used. This presents a highly beneficial use of MR technology to minimize exposure and optimize PPE use for HCWs. Furthermore, Liu et al [41] evaluated the use of MR techniques to improve medical education and understanding of pulmonary lesions resulting from COVID-19 infection. The researchers concluded that the group's mean task score using 3D holograms provided by MR techniques was significantly higher than that of the group using standard 2D computed tomography imaging. Moreover, the group using MR technology scored substantially lower for the mental, temporal performance, and frustration subscales on the National Aeronautics and Space Administration Task Load Index questionnaire. These results highlight the use of MR tools in medical education to improve understandability, spatial awareness, and interest and lower the learning curve. Similarly, Muangpoon et al [42] used MR to support benchtop models for digital rectal examinations to improve visualization and learning. The evaluation of such a MR system showed that the increased visualization allowed for enhanced learning, teaching, and assessment of digital rectal examinations. Hilt et al [43] examined the use of MR technologies to provide patient education on myocardial infarction. The researchers concluded that MR technologies act as a practical tool to unite disease perspectives between patients and professionals as well as optimize knowledge transfer. In addition, House et al [44] investigated the use of an MR tool, *VSI Patient Education*, to provide superior education before epilepsy surgery or stereotactic electrode implantation compared with standard 3D rubber brain models. The results showed that the MR tool provided more comprehensible and imaginable patient education than the rubber brain model. In addition, the patients showed a higher preference for the VSI Patient Education tool, emphasizing the benefits of MR tools as the future for patient education. Overall, the rapid acceleration of MR technologies has supported the accessibility and quality of care while also protecting health care staff [40]. When deploying such technologies, topics such as information security, infection control, user experience, and workflow integration must be considered [40]. Such use cases and related requirements must be incorporated into new policy interventions to ensure maximum impact by MR technologies.

Methods

Overview

In this study, text data sets were extracted from Twitter. Three human annotators rated the tweets on a positive, negative, neutral, and inconsistent scale for different factors. We used an interannotator and the mean of the ratings to agree with all the human annotators. The annotated tweets were converted into numerical data using 4 word-embedding models: bag-of-words, term frequency–inverse document frequency, Word2vec, and Doc2Vec. Then, we divided the data set into training and testing with a 4:1 ratio and further divided into training and validation in the ratio of 7:3. Our choice to split the data set into the following ratios was derived from prior work on sentiment analysis evaluation. Specifically, Khagi et al [45] evaluated the performance classification accuracy with a 7:3 ratio with a 5-fold cross-validation. Furthermore, Singh and Kumari [46] used a 4:1 training to testing ratio for sentiment classification. We used a stratified random sampling technique to split these data. Stratified random sampling divides the entire population into homogeneous groups called strata (plural for stratum). Random

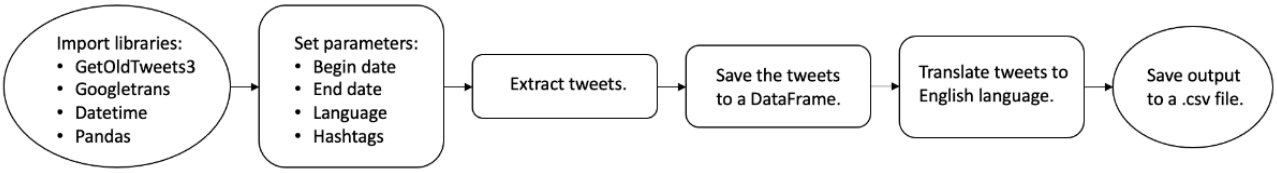
XSL•FO
RenderX

samples were then selected from each stratum. Finally, we used 4 classification models to classify the sentiment of each tweet.

Data Extraction and Preprocessing

The "GetOldTweets3" library from Python was used to extract the tweets. The data corpus consists of tweets posted between November 7, 2019, and August 31, 2020, shortly after the pandemic, which were filtered based on the hashtag, "hololens2," and relevant terms including "holo lens 2" and "hololens 2." We downloaded 8492 tweets, which on average consisted of 20 words each. This study also considered tweets in multiple languages. The corpus contained 5379 tweets in English; 2630 tweets in Japanese; 102 tweets in French; and small portions of German, Spanish, Dutch, and Swedish. A translator from the "googletrans" library in Python was used to translate the tweets into English. Googletrans uses the Google Translate Ajax application programming interface to perform these translations. This translation was performed to enable human annotators to rate the sentiment and improve accuracy rather than machine annotators. The data set did not contain retweets, which would add redundancy to the analysis. Quoted tweets were included if additional texts were included in the search term. Figure 1 shows the flowchart of the data extraction process in Jupyter using Python programming language.

Figure 1. Flowchart of the data extraction process.



After the data collection process, 3 human annotators determined the sentiment of the tweets. Each annotator rated the tweet with respect to the following aspects: usability, field of view, motion sickness, comfort, immersion, cost, and development. This rating was on a scale of positive, negative, neutral, and inconclusive. Positive was rated if the tweet conveyed a positive sentiment toward an attribute. Negative was rated if the tweet conveyed a negative sentiment toward an attribute. Neutral was rated when the tweet did not convey a positive or negative attitude toward an attribute. Finally, inconclusive was rated if the tweet had mixed sentiments or did not have any information related to that specific attribute. Furthermore, human annotators rated the tweets (yes or no) based on the suitability for insights to MR developers, MR user experience researchers, or MR customers and users.

As manually annotating tweets is mostly a subjective process, there were a few instances where the perspective of different annotators was not in agreement. Therefore, to address this challenge, we performed an interannotator agreement. We quantified each positive, negative, neutral, and inconsistent sentiment with a numeric value (ie, 1, −1, 0, and 0). To ease the computation of the interannotator agreement score, the inconsistent label was marked as 0 so that the overall agreement score remained unaffected. The mean of these values was computed using equation (1):



If the mean value was close to −1 and 1, we regarded the annotator perspective as a match. If the mean value was close to 0, we marked that the annotators disagreed with the sentiment conveyed by the tweet. Next, we calculated the average of all the attributes with respect to a tweet to determine the overall sentiment. If this average was positive, we classified the tweet as positive; otherwise, it was classified as negative.

Word-Embedding Models

Bag-of-Words Model

A bag-of-words model represents a method to describe the occurrence of words within a document [47]. It involves two factors: (1) a vocabulary of known words and (2) a measure of the presence of known words. It is referred to as a "bag" of words because the corresponding document is viewed as a set of words rather than a sequence of words. The document's meaning is often well represented by the set of words, whereas the actual word order is ignored. As such, from the content alone, the document's meaning can be determined. Zhang et al [48] developed 2 algorithms that do not rely on clustering and achieved competitive performance in object categorization compared with clustering-based bag-of-words representations. They were successful in achieving better results with their approach. Wu et al [49] proposed a bag-of-words model that mapped semantically related features to the same visual words. Their proposed scheme was effective, and it greatly enhanced the performance of the bag-of-words model.

Term Frequency–Inverse Document Frequency

The term frequency–inverse document frequency (TF-IDF) is a numerical statistic intended to reflect how important a word is to a document in a collection or corpus [50]. It is one of the most widely used techniques for key word detection [51]. The TF-IDF value increases proportionally with the number of times a word appears in the document. However, it is essential to not only consider the number of times a given word occurs in a document but also consider how frequently the word appears in other documents [51]. For example, certain words, referred to as stopwords, such as "is," "of," and "that" frequently appear in documents yet have little importance. To compensate, the TF-IDF value increases with the number of times a word appears in a document but is also offset by the occurrence of that word with a corpus [52]. Peng et al [53] evaluated a novel TF-IDF improved feature weighting approach that reflected the importance of the term among different types of documents. This was achieved by considering the positive or negative set

XSL•FO
RenderX

and weighing the term appropriately. This study showed that the term frequency–inverse positive-negative document frequency classifier outperforms the standard TF-IDF technique. In addition, the results of this study highlight the importance of this analysis technique for imbalanced data sets, which, if not accounted for, could lead to erroneous results [54].

Word2vec

Word2vec is a combination of models, the continuous bag-of-words and skip-gram, used to represent distributed representations of words in a corpus C [55]. Word2vec is an algorithm that accepts a text corpus as an input and outputs a vector representation for each word [56]. Word2vec outputs word vectors that can be represented as a large piece of text or even the entire article [57]. Unlike most test classification techniques, Word2vec uses both a supervised and unsupervised approach. In particular, it is supervised as the model derives a supervised learning task using continuous bags or words and a skip-gram. Furthermore, it is unsupervised, given that any large corpus of choice can be provided [58]. Word2vec cannot determine the importance of each word within a document; therefore, it is challenging to extract which words hold higher importance, comparatively [58]. Ma et al [59] applied the Word2vec technique in big data processing to cluster similar data and reduce the dimension. The results showed that training data fed into Word2vec decreased the data dimension and sped up multiclass classification. Lilleberg et al [58] found that a combination of Word2vec and TF-IDF outperformed TF-IDF.

Doc2Vec

Doc2Vec also uses an unsupervised learning approach to learn document representation [60]. It can be used to identify abnormal comments and recommend relevant topics to users [61,62]. The input of texts (ie, words) per document can be varied, whereas the output is a fixed-length vector [59]. It is a modified version of the Word2vec algorithm using paragraph vectors [63]. Paragraph vectors are unique among all documents, whereas word vectors are shared among all documents. Word vectors can be learned from different documents. Word vectors will be trained during the training phase, while paragraphs will be thrown away after that. During the prediction phase, paragraph vectors will be initialized randomly and computed using word vectors. The main difference between Doc2Vec and Word2Vec is that the latter computes a vector for every word in the document, whereas Doc2Vec computes a vector for the entire document in the corpus. Using Word2Vec and Doc2Vec together will yield significantly better results and promote a thorough study of any document.

Classification Models

Logistic Regression

The logistic regression model is based on the odds of the binary outcomes of interest [64]. For simplicity, one outcome level is designated as the event of interest. In the following text, it is simply called the event. The odds of the event are the ratio of the probability of the event occurring divided by the likelihood of the event not occurring. Odds are often used for gambling, and “even odds” (odds=1) correspond to the event happening half the time. This would be the case for rolling an even number on a single die. The odds for rolling a number <5 would be 2 because rolling a number <5 is twice as likely as rolling a number 5 or 6. Symmetry in the odds is found by taking the reciprocal. The odds of rolling at least a 5 would be 0.5 (=1/2). The logistic regression model takes the natural logarithm of the odds as a regression function of the predictors. With 1 predictor, X, this takes the form $\ln[\text{odds}(Y=1)]=\beta 0+\beta 1X$, where ln stands for the natural logarithm, Y is the outcome, where Y=1 occurs when the event occurs and Y=0 when it does not, β0 is the intercept term, and β1 represents the regression coefficient, the change in the logarithm of the event odds with a 1-unit change in the predictor X. The difference in the logarithms of 2 values is equal to the logarithm of the ratio of the 2 values. Thus, by taking the exponential of β1, we obtain the odds ratio corresponding to a 1-unit change in X. The logistic regression model has been used in many social media–based sentiment analysis studies [65-67].

Random Forest

Random forest is an ensemble learning method based on the decision tree algorithm [68]. It uses multiple decision trees and merges them to provide absolute and stable outcomes, mostly used for training and class output. Many previous studies successfully used the decision tree and random forest algorithms for sentiment classification of social media data [69-72].

XGBoost

The XGBoost (eXtreme Gradient Boos) is a scalable end-to-end tree boosting system for tree boosting, which uses a sparsity aware algorithm to handle sparse data sets [73]. Although the XGBoost uses a representation similar to that of random forest, the prediction error is significantly lower than that of the random forest. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models, which are then added together to make the final prediction. It is called gradient boosting, as it uses a gradient descent algorithm to minimize the loss when adding new models. The gradient boosting algorithm achieves results faster and performs efficiently compared with other algorithms. Aziz and Dimililer [74] used an ensemble XGBoost classifier to enhance sentiment analysis in social media data and demonstrated an improvement of the sentiment classification performance.

Support Vector Machines

A support vector machine (SVM) is a supervised learning model for 2-group classification problems by locating a hyperplane in a multidimensional space that clearly separates the data points [75,76]. The main purpose of SVM is to determine an optimal separating hyperplane that not only separates the data but also ensures that the margin to the data on both sides is as large as possible. First, an optimal solution in a low-dimensional space that can aptly separate the data is evaluated. If this is not possible, the data are mapped to a high-dimensional space by using nonlinear transformation methods. From this, a valid kernel function is selected to determine the optimal linear classification surface. It is highly efficient in separating data into different classes. This allows us to group words into different categories, which helps us access the words easily.

XSL•FO
RenderX

The SVM model has been used in various sentiment analysis studies and has produced high classification accuracy [77-79].

Ethics Approval

This research does not require institutional review board approval because the project does not include any interaction or intervention with human subjects.

Results

Model Learning and Performance

Once we determined the classified sentiment for each tweet, we trained a model for sentiment analysis using supervised learning. First, we evaluated the imbalance in the data set: 527 positive tweets and 229 negative tweets. We collected data from 516 unique users in this study. The minimum number of tweets per user was 1, whereas the maximum was 18. The average number of tweets per user was 1.50 (SD 0.3).

To perform supervised learning, it was necessary to preprocess the data. We cleaned the data by removing punctuations, stop words, single characters, and uneven spaces; converting the data to lower case; and stemming on these data. Following preprocessing, we tokenized the data using 4 different techniques: bag-of-words, TF-IDF, Word2vec, and Doc2Vec. Table 2 lists the performance of each model with different word embeddings over a training test ratio of 80:20. This table shows that the bag-of-words tokenizing method using a random forest supervised learning approach produced the highest accuracy of the test set at 81.29%. Furthermore, Textbox 1 summarizes the top words that contribute toward sentiment classification. This textbox highlights various words contributing to sentiments, such as “problem,” “mess,” and “error” for negative and “nice,” “love,” and “achieve” for positive.

Table 2. The performance percentage of each model with different work embeddings.

Method and set	Logistic regression	Random forest	XGBoost	SVM[a]
Bag-of-words				
Validation	69.18	72.79	65.40	69.72
Test	69.03	81.29	72.25	75.48
TF-IDF[b]				
Validation	69.72	74.05	62.16	70.81
Test	74.83	76.12	74.83	78.70
Word2vec				
Validation	65.40	68.10	68.84	67.02
Test	72.25	74.83	77.41	71.61
Doc2Vec				
Validation	66.48	67.56	66.48	68.10
Test	70.32	67.74	70.32	69.03

[a]SVM: support vector machine.

[b]TF-IDF: term frequency–inverse document frequency.

XSL•FO
RenderX

Textbox 1. Most significant words used in the sentiment analysis.

- mrdevdays
- talking
- thinking
- pc
- mvis
- azure
- mess
- market
- think
- announced
- use
- knowledge
- yotiky
- markets
- lightning
- firefox
- achieve
- babylon
- hatenablog
- nice
- playing
- july
- emulator
- available
- hololens2
- microvision
- love
- today
- general
- keynote
- mxdrealitydev
- hololens
- snapchat
- terrible
- solve
- problem
- forehead
- time
- buy
- msdevirl
- probably
- million
- altspacevr

XSL•FO
RenderX

- microsoft
- half
- nreal
- procedure
- error
- office

Insights From the Perspective of the COVID-19 Pandemic and Health Care

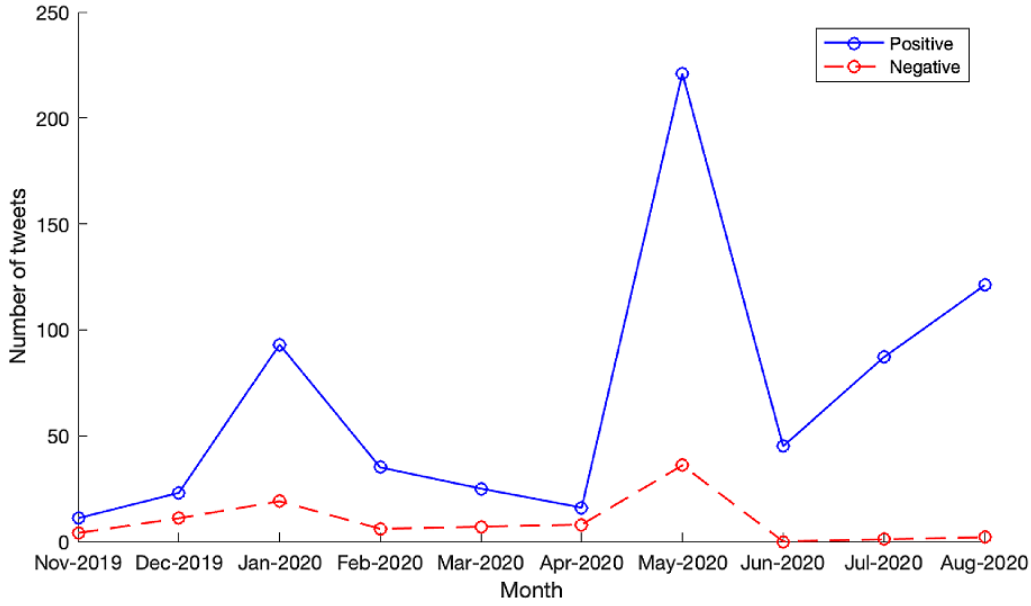
Following the determination of an appropriate classification model, we evaluated the reasoning for positive or negative tweet classification. Upon investigation, words like "COVID," "pandemic," "patients," and "health care" were all associated with the positive sentiment. Further evaluation showed that the use of HoloLens 2 is highly encouraged in the health care industry in several respects. First, tweets showed the use of HoloLens2 to enable virtual appointments in times of unprecedented crisis. As such, HCW found HoloLens2 to be a vital tool to improve safety and quality of care while also being easy to set up and comfortable to wear. This finding is significant as it supports previous studies evaluating the capabilities of MR technology to permit telemedicine [40]. Other tweets highlighted the use of HoloLens2 to facilitate education and training during the pandemic. Specifically, the HoloLens2 enabled HCW to practice coronavirus identification in a socially distanced manner, which minimized the risk of contact and transmission. Similarly, this finding is significant as it supports prior works relating to the use of MR tools to improve medical education and understanding [41-43]. The following are examples of tweets that qualitatively support these insights:

1. We are revolutionizing healthcare using @Microsoft #HoloLens2 to deliver remote care in #COVID19! Staff found it easy to set up, comfortable to wear, improved quality of care. #Hololens2 is helping keep our #healthcareworkers stay safe on the frontline!
2. "Use of #HoloLens Mixed Reality Headset for Protecting Health Care Workers During the #COVID19 Pandemic": Prospective study used @Microsoft HoloLens2 to support remote patient care for hospitalized patients. Reduced exposure time by 51% & PPE usage by 83%:
3. Nowadays #medical industry getting lots of advancement with recent tech. There are many notable advantages of #Microsoft #HoloLens that prove that the future of #healthcare is heavily reliant on #MixedReality technology #MR #XR #Hololens2 #AR #Remote
4. #HoloLens2 helps safely train doctors to identify #coronavirus in patients. #MixedReality offers the perfect, socially-distanced or remote training experience, minimizing contact, risk and transmission.
5. Use of the HoloLens2 Mixed Reality Headset for Protecting Health Care Workers During the COVID-19 Pandemic: Prospective, Observational Evaluation

Changes in sentiment toward HoloLens2 throughout the pandemic were also evaluated. In November 2019, when HoloLens 2 was released, there was no significant difference in the positive and negative sentiment (Figure 2). This is likely caused by consumer delay to learn about the product's arrival in the market supported by the low tweet volumes of both positive and negative sentiments. In December 2019 and January 2020, a significant increase in the positive view was observed, likely caused by consumer interest in the newly released product. In February 2020, the onset of the pandemic occurred, which resulted in the severely affected sales of consumer goods. This period aligns with the drop in general sentiment on both sides. However, the general sentiment of HoloLens 2 seems to be positive despite affected sales. In May 2020, there was a sudden increase in positive sentiment. It is hypothesized that consumers, especially in health care, noticed the device's benefits to minimize the risk of contraction and transmission. Following this significant change in sentiment, the negative sentiment toward the device almost dropped to 0, highlighting the continued positive role of HoloLens2 during the pandemic.

XSL•FO
RenderX

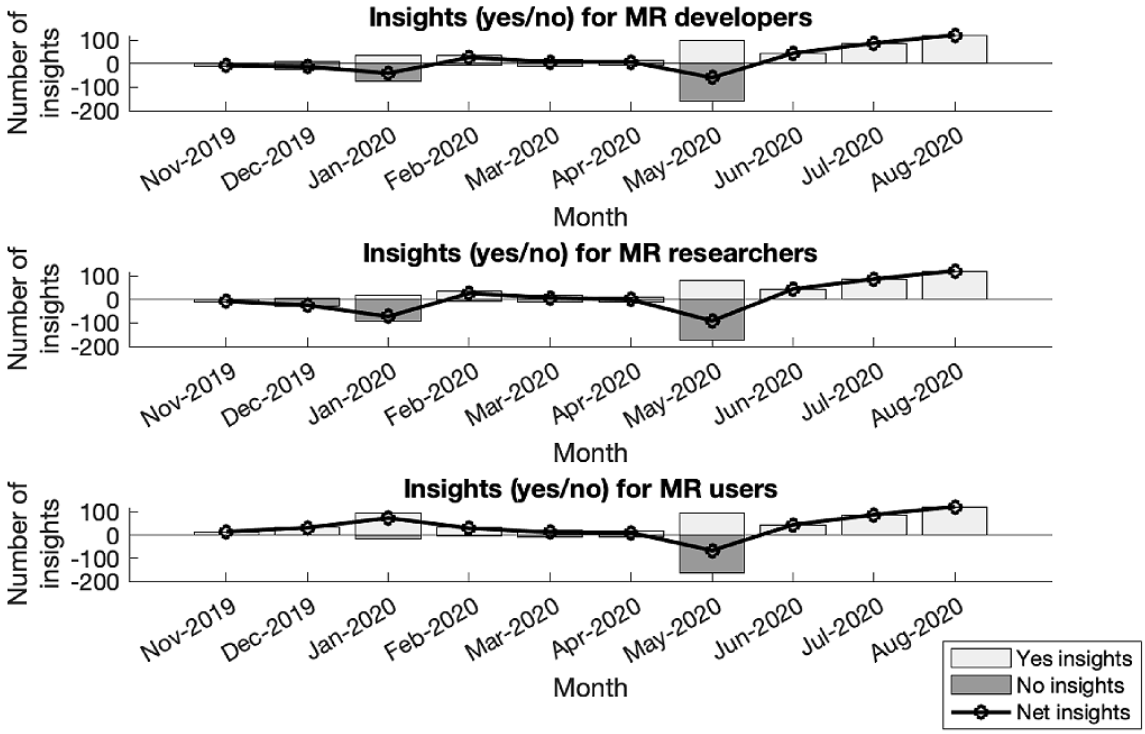
Figure 2. Tweet sentiment over time.



Insights for MR Developers, Researchers, and Users

Figure 3 breaks down the tweets into useful insights for MR developers, defined as individuals developing features of the technology, researchers, defined as individuals using the device for research endeavors (ie, usability analyses), and users, defined as individuals using the device for leisure. The green bars represent tweets classified as suitable insights, and red bars as not suitable. Furthermore, we calculated the net insights, indicated by the black line, as the suitable insights (yes) minus the not suitable insights (no). In the first few months, the data are distributed equally on both sides, and the net insight is approximately 0. In May 2020, there is a drastic difference in the distribution. We presume that this sudden charge is because of the largely changing technology uses caused by the pandemic. Following, we predict that the steady increase in suitable insights results from individuals becoming more acclimated to the technology-driven, remote lifestyle.

Figure 3. Suitability of tweets to provide insights to mixed reality (MR) developers, researchers, and users after its release.



XSL•FO
RenderX

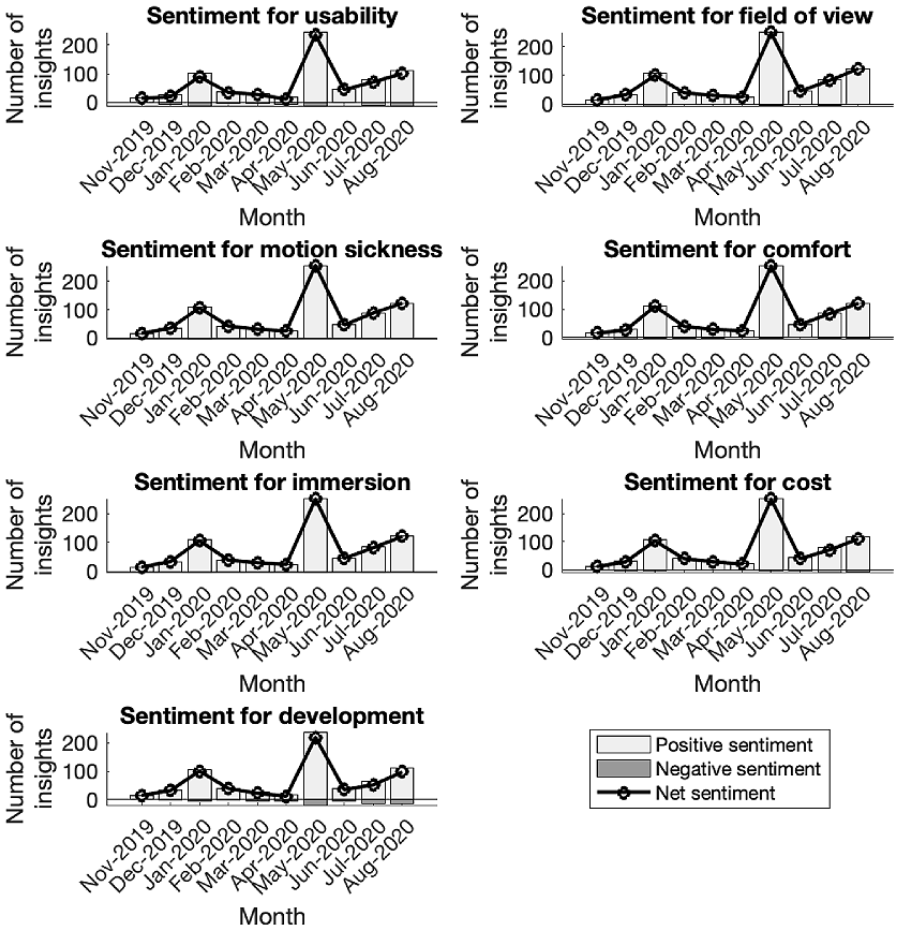
Analysis of HoloLens2 Characteristics

Table 3 shows the net sentiment of various factors related to HoloLens 2 over the analyzed period. Furthermore, Figure 4 illustrates the number of positive sentiments as green bars, negative sentiments as red bars, and net sentiment as the black line for all factors. We calculated the net sentiment as the number of tweets with positive sentiment minus the number of tweets with negative sentiment. The results show that net sentiment is exclusively positive for all factors in all the months studied. It shows a positive trend in usability, field of view, motion sickness, comfort, immersion, cost, and development. All these factors contributed to positive sentiment toward HoloLens 2. This trend can be credited to the impact of the COVID-19 pandemic as the number of people depending on this device increased.

Table 3. Stacked net sentiments related to various factors over 10 months.

Month	Usability	Field of view	Motion sickness	Comfort	Immersion	Cost	Development
November 19	+15	+15	+15	+15	+15	+11	+15
December 19	+20	+32	+34	+26	+34	+28	+32
January 20	+90	+100	+106	+110	+106	+102	+100
February 20	+35	+39	+41	+39	+39	+39	+39
March 20	+28	+30	+30	+28	+30	+28	+22
April 20	+12	+24	+24	+24	+24	+20	+12
May 20	+235	+249	+257	+253	+251	+251	+219
June 20	+45	+45	+45	+45	+45	+39	+35
July 20	+72	+82	+88	+84	+82	+70	+52
August 20	+101	+123	+123	+121	+121	+109	+97

Figure 4. Positive, negative, and net sentiments related to various factors over 10 months.



XSL•FO
RenderX

Discussion

Principal Findings

The bag-of-words tokenizing method, using a random forest supervised learning approach, provided the highest accuracy of the test set at 81.29%, according to the results of our sentiment analysis. Furthermore, the findings reveal an apparent shift in public opinion during the pandemic. Consumer products were significantly affected during the pandemic's start, which coincided with a dip in both positive and negative emotion. Following that, there is a sharp increase in positive feeling, which is thought to be because of the device's new applications in health care teaching and training. This coincides with significant shifts in the number of practical insights for MR developers, researchers, and users, as well as positive net attitudes for HoloLens 2 features.

Twitter is one of the most popular social media platforms worldwide. In this study, tweets related to HoloLens 2 were obtained; however, they did not cover all opinions. We only used tweets with the hashtag "hololens2." Therefore, many tweets related to this topic, without the hashtag, might have been left out. In addition, this resulted in a relatively small sample size comparatively. Furthermore, some individuals might use other platforms to state their opinion about a particular device. For example, some individuals tend to make reviews or first-opinion videos of devices on platforms such as YouTube, which generate much discussion in the comments. These comments also contribute to consumer perceptions of the product. In addition, we could have explored other social media platforms, such as Instagram and Facebook. The literature supports the use of YouTube, Instagram, and Facebook for sentiment analysis. For example, sentiment analysis has been studied to determine the most relevant and popular video on YouTube according to the search [80]. Furthermore, a deep neural network can be used to propose a sentiment analysis model of YouTube comments [81]. Other researchers used a sentiment analysis tool to measure the proposed social value of each image [82]. Ortigosa et al [83] stated that adaptive e-learning systems could use sentiment analysis to support personalized learning. Adding additional platforms in this study would contribute to a greater understanding of consumer perception. Finally, the extent to which the data were sampled may introduce some biases. Less than half of the adults regularly use Twitter; individuals between the ages of 18 and 29 years as well as minorities are highly represented on Twitter compared with the general population, and Twitter consists of almost entirely passive users (<50 tweets per year) and very active users (>1000 tweets per year) [84]. Therefore, these limitations may have resulted in certain samples of the population being more represented than others.

The onset of the pandemic occurred from February 2020. During the first couple of months, we observed a sudden increase in the popularity of HoloLens 2, which was primarily attributed to new use cases in the health care field. In addition, this change can likely be credited to the large shift to working or studying from home. This analysis covered only a portion of the pandemic when the world began adapting to new routines, technologies, and lifestyles. It would have been beneficial to include tweets made a couple of months after August 2020, as this was the period when people were more adapted to working and studying from home. Including more months would provide increased insight on user sentiment over time through the pandemic, enabling a more thorough understanding.

Conclusions

In this study, we used aspect-based sentiment analysis to study the usability of HoloLens 2. We extracted data from Twitter based on the hashtag "hololens2" to explore user perception about HoloLens 2. We accumulated 8492 tweets and translated the non-English tweets into English using the "googletrans" library in Python. After the data collection process, human annotators rated the tweets on a positive, negative, neutral, and inconsistent scale for 7 different factors and determined the suitability of the tweets to provide insights for MR developers, researchers, and users. We used an interannotator and rating average to ensure agreement among the human annotators. The results show a clear indication between the positive and negative sentiments toward HoloLens 2. Specifically, we observed that the positive sentiment toward the device grew during the onset of the COVID-19 pandemic, whereas the negative sentiment decreased. By separating the most popular words from both sentiments, we identified the positive and negative aspects of the device. We also observed that HoloLens 2 was highly encouraged in the health care industry. A close evaluation of tweets found that HoloLens 2 enabled virtual appointments, supported medical training, and provided patient education. As such, this thematic analysis showed that HoloLens 2 facilitated social distance practices, which largely minimized the risk of contraction and transmission. The findings of this study contribute to a more holistic understanding of public perception and acceptance of VR and AR technologies, especially during the unprecedented COVID-19 pandemic. Further, these findings highlight several new implementations of HoloLens 2 in health care, which may inspire future use cases. In future work, more data from various social media platforms will be included and compared to improve the effectiveness of this process.

Conflicts of Interest

None declared.

References

1. Martinez-Millana A, Bayo-Monton J, Lizondo A, Fernandez-Llatas C, Traver V. Evaluation of Google glass technical limitations on their integration in medical systems. Sensors (Basel) 2016 Dec 15;16(12):2142 [FREE Full text] [doi: 10.3390/s16122142] [Medline: 27983691]

XSL•FO
RenderX

2. Broach J, Hart A, Griswold M, Lai J, Boyer EW, Skolnik AB, et al. Usability and reliability of smart glasses for secondary triage during mass casualty incidents. Proc Annu Hawaii Int Conf Syst Sci 2018 Jan 03;2018:1416-1422 [FREE Full text] [Medline: 29398976]
3. O'Hagan J, Khamis M, Williamson JR. Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '21). 2021 Presented at: MMSys '21: 12th ACM Multimedia Systems Conference; Sep 28-Oct 1, 2021; Istanbul Turkey. [doi: 10.1145/3458307.3460965]
4. Dey A, Billinghurst M, Lindeman RW, Swan JE. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Front Robot AI 2018 Apr 17;5:37 [FREE Full text] [doi: 10.3389/frobt.2018.00037] [Medline: 33500923]
5. Pereira R, Moore HF, Gheisari M, Esmaeili B. Development and usability testing of a panoramic augmented reality environment for fall hazard safety training. In: Advances in Informatics and Computing in Civil and Construction Engineering. Cham: Springer; 2018.
6. Vinci C, Brandon KO, Kleinjan M, Hernandez LM, Sawyer LE, Haneke J, et al. Augmented reality for smoking cessation: development and usability study. JMIR Mhealth Uhealth 2020 Dec 31;8(12):e21643 [FREE Full text] [doi: 10.2196/21643] [Medline: 33382377]
7. HoloLens 2 vs HoloLens 1: what's new? 4Experience Virtual Reality Studio. URL: https://4experience.co/hololens-2-vs-hololens-1-whats-new/ [accessed 2022-01-25]
8. Geszten D, Komlódi A, Hercegfi K, Hámornik B, Young A, Köles M, et al. A content-analysis approach for exploring usability problems in a collaborative virtual environment. Acta Polytechnica Hungarica 2018 Nov 06;15(5):67. [doi: 10.12700/aph.15.5.2018.5.5]
9. Carroll WM. Emerging Technologies for Nurses Implications for Practice. Cham: Springer; 2020.
10. Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M. Augmented reality technologies, systems and applications. Multimed Tools Appl 2010 Dec 14;51(1):341-377. [doi: 10.1007/s11042-010-0660-6]
11. Egliston B, Carter M. Oculus imaginaries: the promises and perils of Facebook's virtual reality. New Media Soc 2020 Sep 24;24(1):70-89. [doi: 10.1177/1461444820960411]
12. Yildirim C, Carroll M, Hufnal D, Johnson T, Pericles S. Video game user experience: to VR, or not to VR? In: Proceedings of the 2018 IEEE Games, Entertainment, Media Conference (GEM). 2018 Presented at: 2018 IEEE Games, Entertainment, Media Conference (GEM); Aug 15-17, 2018; Galway, Ireland. [doi: 10.1109/gem.2018.8516542]
13. Bayro A, Ghasemi Y, Jeong H. Subjective and objective analyses of collaboration and co-presence in a virtual reality remote environment. In: Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2022 Presented at: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); Mar 12-16, 2022; Christchurch, New Zealand. [doi: 10.1109/vrw55335.2022.00108]
14. Wei NJ, Dougherty B, Myers A, Badawy SM. Using google glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
15. Rauschnabel PA, Brem A, Ivens BS. Who will buy smart glasses? Empirical results of two pre-market-entry studies on the role of personality in individual awareness and intended adoption of Google Glass wearables. Comput Human Behav 2015 Aug;49:635-647. [doi: 10.1016/j.chb.2015.03.003]
16. Ghasemi Y, Jeong H. Model-based task analysis and large-scale video-based remote evaluation methods for extended reality research. arXiv. Preprint posted online on March 13, 2021. [FREE Full text]
17. Hammady R, Ma M, Strathearn C. User experience design for mixed reality: a case study of HoloLens in museum. Int J Technol Market 2019;13(3/4):354. [doi: 10.1504/ijtmkt.2019.104600]
18. Hoover M, Miller J, Gilbert S, Winer E. Measuring the performance impact of using the Microsoft HoloLens 1 to provide guided assembly work instructions. J Comput Inf Sci Eng 2020 Dec;20(6):061001. [doi: 10.1115/1.4046006]
19. Xue H, Sharma P, Wild F. User satisfaction in augmented reality-based training using Microsoft HoloLens. Computers 2019 Jan 25;8(1):9. [doi: 10.3390/computers8010009]
20. AR in order-picking – experimental evidence with Microsoft HoloLens. Mensch und Computer 2018 - Workshopband. URL: https://dl.gi.de/handle/20.500.12116/16900 [accessed 2022-01-25]
21. Levy JB, Kong E, Johnson N, Khetarpal A, Tomlinson J, Martin GF, et al. The mixed reality medical ward round with the MS HoloLens 2: innovation in reducing COVID-19 transmission and PPE usage. Future Healthc J 2021 Mar 18;8(1):e127-e130 [FREE Full text] [doi: 10.7861/fhj.2020-0146] [Medline: 33791491]
22. Park BJ, Hunt SJ, Nadolski GJ, Gade TP. Augmented reality improves procedural efficiency and reduces radiation dose for CT-guided lesion targeting: a phantom study using HoloLens 2. Sci Rep 2020 Oct 29;10(1):18620 [FREE Full text] [doi: 10.1038/s41598-020-75676-4] [Medline: 33122766]
23. Thees M, Kapp S, Strzys MP, Beil F, Lukowicz P, Kuhn J. Effects of augmented reality on learning and cognitive load in university physics laboratory courses. Comput Human Behav 2020 Jul;108:106316. [doi: 10.1016/j.chb.2020.106316]
24. Kharde AV, Sonawane S. Sentiment analysis of twitter data: a survey of techniques. Int J Comput Application 2016 Apr 15;139(11):5-15. [doi: 10.5120/ijca2016908625]
25. Carvalho J, Plastino A. On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. Artif Intell Rev 2020 Aug 27;54(3):1887-1936. [doi: 10.1007/s10462-020-09895-6]

XSL•FO
RenderX

26. Guo X, Li J. A novel Twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency. In: Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). 2019 Presented at: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS); Oct 22-25, 2019; Granada, Spain. [doi: 10.1109/snams.2019.8931720]
27. Chamlertwat W, Bhattarakosol P, Rungkasiri T, Haruechaiyasak C. Discovering consumer insight from Twitter via sentiment analysis. J Universal Comput Sci 2012;18(8):973-992. [doi: 10.3217/jucs-018-08-0973]
28. Bian J, Yoshigoe K, Hicks A, Yuan J, He Z, Xie M, et al. Mining Twitter to assess the public perception of the "Internet of Things". PLoS One 2016;11(7):e0158450 [FREE Full text] [doi: 10.1371/journal.pone.0158450] [Medline: 27391760]
29. Mittal A, Goel A. Stock prediction using twitter sentiment analysis. Stanford. URL: http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf [accessed 2022-01-25]
30. Venugopalan M, Gupta D. Exploring sentiment analysis on twitter data. In: Proceedings of the 2015 Eighth International Conference on Contemporary Computing (IC3). 2015 Presented at: 2015 Eighth International Conference on Contemporary Computing (IC3); Aug 20-22, 2015; Noida, India. [doi: 10.1109/ic3.2015.7346686]
31. Troisi O, Grimaldi M, Loia F, Maione G. Big data and sentiment analysis to highlight decision behaviours: a case study for student population. Behav Inform Technol 2018 Jul 23;37(10-11):1111-1128. [doi: 10.1080/0144929x.2018.1502355]
32. Nanath K, Joy G. Leveraging Twitter data to analyze the virality of Covid-19 tweets: a text mining approach. Behav Inform Technol 2021 Jun 17:1-19. [doi: 10.1080/0144929x.2021.1941259]
33. Nguyen TT, Meng H, Sandeep S, McCullough M, Yu W, Lau Y, et al. Twitter-derived measures of sentiment towards minorities (2015-2016) and associations with low birth weight and preterm birth in the United States. Comput Human Behav 2018 Dec;89:308-315 [FREE Full text] [doi: 10.1016/j.chb.2018.08.010] [Medline: 30923420]
34. Gaspar R, Pedro C, Panagiotopoulos P, Seibt B. Beyond positive or negative: qualitative sentiment analysis of social media reactions to unexpected stressful events. Comput Human Behav 2016 Mar;56:179-191. [doi: 10.1016/j.chb.2015.11.040]
35. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
36. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
37. Berkovic D, Ackerman IN, Briggs AM, Ayton D. Tweets by people with arthritis during the COVID-19 pandemic: content and sentiment analysis. J Med Internet Res 2020 Dec 03;22(12):e24550 [FREE Full text] [doi: 10.2196/24550] [Medline: 33170802]
38. Shahzad K, Malik MK, Mehmood K. Perception of wearable intelligent devices: a case of fitbit-alta-HR. In: Intelligent Technologies and Applications. Singapore: Springer; 2020.
39. El-Gayar O, Nasralah T, Noshokaty AE. Wearable devices for health and wellbeing: design Insights from Twitter. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; 2019; Hawaii, USA. [doi: 10.24251/hicss.2019.467]
40. Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, PanSurg Collaborative. Use of the HoloLens2 mixed reality headset for protecting health care workers during the COVID-19 pandemic: prospective, observational evaluation. J Med Internet Res 2020 Aug 14;22(8):e21486 [FREE Full text] [doi: 10.2196/21486] [Medline: 32730222]
41. Liu S, Xie M, Zhang Z, Wu X, Gao F, Lu L, et al. A 3D hologram with mixed reality techniques to improve understanding of pulmonary lesions caused by COVID-19: randomized controlled trial. J Med Internet Res 2021 Sep 10;23(9):e24081 [FREE Full text] [doi: 10.2196/24081] [Medline: 34061760]
42. Muangpoon T, Haghighi Osgouei R, Escobar-Castillejos D, Kontovounisios C, Bello F. Augmented reality system for digital rectal examination training and assessment: system validation. J Med Internet Res 2020 Aug 13;22(8):e18637 [FREE Full text] [doi: 10.2196/18637] [Medline: 32788146]
43. Hilt AD, Mamaqi Kapllani K, Hierck BP, Kemp AC, Albayrak A, Melles M, et al. Perspectives of patients and professionals on information and education after myocardial infarction with insight for mixed reality implementation: cross-sectional interview study. JMIR Hum Factors 2020 Jun 23;7(2):e17147 [FREE Full text] [doi: 10.2196/17147] [Medline: 32573464]
44. House PM, Pelzl S, Furrer S, Lanz M, Simova O, Voges B, et al. Use of the mixed reality tool "VSI Patient Education" for more comprehensible and imaginable patient educations before epilepsy surgery and stereotactic implantation of DBS or stereo-EEG electrodes. Epilepsy Res 2020 Jan;159:106247. [doi: 10.1016/j.eplepsyres.2019.106247] [Medline: 31794952]
45. Khagi B, Kwon G, Lama R. Comparative analysis of Alzheimer's disease classification by CDR level using CNN, feature selection, and machine‐learning techniques. Int J Imaging Syst Technol 2019 Mar 07;29(3):297-310. [doi: 10.1002/ima.22316]
46. Singh T, Kumari M. Role of text pre-processing in twitter sentiment analysis. Procedia Comput Sci 2016;89:549-554. [doi: 10.1016/j.procs.2016.06.095]
47. Brownlee J. A gentle introduction to the bag-of-words model. Deep Learning for Natural Language Processing. 2017. URL: https://machinelearningmastery.com/gentle-introduction-bag-words-model/ [accessed 2022-01-25]

XSL•FO
RenderX

48. Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. Int J Mach Learn Cyber 2010 Aug 28;1(1-4):43-52. [doi: 10.1007/s13042-010-0001-0]
49. Lei W, Hoi SC, Nenghai Y. Semantics-preserving bag-of-words models and applications. IEEE Trans Image Process 2010 Jul;19(7):1908-1920. [doi: 10.1109/tip.2010.2045169]
50. Singh P. Fundamentals of Bag of Words and TF-IDF. Medium. URL: https://medium.com/analytics-vidhya/fundamentals-of-bag-of-words-and-tf-idf-9846d301ff22 [accessed 2022-01-25]
51. Havrlant L, Kreinovich V. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). Int J Gen Syst 2017 Mar 14;46(1):27-36. [doi: 10.1080/03081079.2017.1291635]
52. Christian H, Agus MP, Suhartono D. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). ComTech 2016 Dec 31;7(4):285. [doi: 10.21512/comtech.v7i4.3746]
53. Peng T, Liu L, Zuo W. PU text classification enhanced by term frequency-inverse document frequency-improved weighting. Concurrency Computat Pract Exper 2013 May 10;26(3):728-741. [doi: 10.1002/cpe.3040]
54. Alshamsi A, Bayari R, Salloum S. Sentiment analysis in English texts. Adv Sci Technol Eng Syst J 2020 Dec;5(6):1683-1689. [doi: 10.25046/aj0506200]
55. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv 2013.
56. Ali Z. A simple Word2vec tutorial. Medium. URL: https://medium.com/@zafaralibagh6/a-simple-word2vec-tutorial-61e64e38a6a1 [accessed 2022-01-25]
57. Ma L, Zhang Y. Using Word2Vec to process big text data. In: Proceedings of the 2015 IEEE International Conference on Big Data (Big Data). 2015 Presented at: 2015 IEEE International Conference on Big Data (Big Data); Oct 29-Nov 01, 2015; Santa Clara, CA, USA. [doi: 10.1109/bigdata.2015.7364114]
58. Lilleberg J, Zhu Y, Zhang Y. Support vector machines and Word2vec for text classification with semantic features. In: Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC). 2015 Presented at: 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC); Jul 06-08, 2015; Beijing, China. [doi: 10.1109/icci-cc.2015.7259377]
59. Ma E. Understand how to transfer your paragraph to vector by doc2vec. Towards Data Science. URL: https://towardsdatascience.com/understand-how-to-transfer-your-paragraph-to-vector-by-doc2vec-1e225ccf102 [accessed 2022-01-25]
60. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning. 2014 Presented at: 31st International Conference on Machine Learning; Jun 22-24, 2014; Bejing, China.
61. Chang W, Xu Z, Zhou S, Cao W. Research on detection methods based on Doc2vec abnormal comments. Future Generation Comput Sys 2018 Sep;86:656-662. [doi: 10.1016/j.future.2018.04.059]
62. Karvelis P, Gavrilis D, Georgoulas G, Stylios C. Topic recommendation using Doc2Vec. In: Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN). 2018 Presented at: 2018 International Joint Conference on Neural Networks (IJCNN); Jul 08-13, 2018; Rio de Janeiro, Brazil. [doi: 10.1109/ijcnn.2018.8489513]
63. Bilgin M, Şentürk IF. Sentiment analysis on Twitter data with semi-supervised Doc2Vec. In: Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK). 2017 Presented at: 2017 International Conference on Computer Science and Engineering (UBMK); Oct 05-08, 2017; Antalya, Turkey. [doi: 10.1109/ubmk.2017.8093492]
64. LaValley MP. Logistic regression. Circulation 2008 May 06;117(18):2395-2399. [doi: 10.1161/circulationaha.106.682658]
65. Bhargava K, Katarya R. An improved lexicon using logistic regression for sentiment analysis. In: Proceedings of the 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN). 2017 Presented at: 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN); Oct 12-14, 2017; Gurgaon, India. [doi: 10.1109/ic3tsn.2017.8284501]
66. Omari MA, Al-Hajj M, Hammami N, Sabra A. Sentiment classifier: logistic regression for Arabic services' reviews in Lebanon. In: Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS). 2019 Presented at: 2019 International Conference on Computer and Information Sciences (ICCIS); Apr 03-04, 2019; Sakaka, Saudi Arabia. [doi: 10.1109/iccisci.2019.8716394]
67. Wasi N, Abulaish M. Document-level sentiment analysis through incorporating prior domain knowledge into logistic regression. In: Proceedings of the 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). 2020 Presented at: 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); Dec 14-17, 2020; Melbourne, Australia. [doi: 10.1109/wiiat50758.2020.00148]
68. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]
69. Karthika P, Murugeswari R, Manoranjithem R. Sentiment analysis of social media network using random forest algorithm. In: Proceedings of the 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). 2019 Presented at: 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS); Apr 11-13, 2019; Tamilnadu, India. [doi: 10.1109/incos45849.2019.8951367]
70. Aufar M, Andreswari R, Pramesti D. Sentiment analysis on YouTube social media using decision tree and random forest algorithm: a case study. In: Proceedings of the 2020 International Conference on Data Science and Its Applications (ICoDSA).

XSL•FO
RenderX

2020 Presented at: 2020 International Conference on Data Science and Its Applications (ICoDSA); Aug 05-06, 2020; Bandung, Indonesia. [doi: 10.1109/icodsa50139.2020.9213078]
71. Singh NK, Tomar DS, Sangaiah AK. Sentiment analysis: a review and comparative analysis over social media. J Ambient Intell Human Comput 2018 May 23;11(1):97-117. [doi: 10.1007/s12652-018-0862-8]
72. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS One 2021 Feb 25;16(2):e0245909 [FREE Full text] [doi: 10.1371/journal.pone.0245909] [Medline: 33630869]
73. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco California USA. [doi: 10.1145/2939672.2939785]
74. Hama Aziz RH, Dimililer N. SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier. J Chinese Institute Eng 2021 Jun 28;44(6):562-572. [doi: 10.1080/02533839.2021.1933598]
75. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998 Jul 10;13(4):18-28. [doi: 10.1109/5254.708428]
76. Noble WS. What is a support vector machine? Nat Biotechnol 2006 Dec;24(12):1565-1567. [doi: 10.1038/nbt1206-1565] [Medline: 17160063]
77. Esparza GG, de-Luna A, Zezzatti AO, Hernandez A, Ponce J, Álvarez M, et al. A sentiment analysis model to analyze students reviews of teacher performance using support vector machines. In: Distributed Computing and Artificial Intelligence, 14th International Conference. Cham: Springer; 2018.
78. Shuai Q, Huang Y, Jin L, Pang L. Sentiment analysis on Chinese hotel reviews with Doc2Vec and classifiers. In: Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). 2018 Presented at: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC); Oct 12-14, 2018; Chongqing, China. [doi: 10.1109/iaeac.2018.8577581]
79. Xia H, Yang Y, Pan X, Zhang Z, An W. Sentiment analysis for online reviews using conditional random fields and support vector machines. Electron Commer Res 2019 May 13;20(2):343-360. [doi: 10.1007/s10660-019-09354-7]
80. Bhuiyan H, Ara J, Bardhan R, Islam MR. Retrieving YouTube video by sentiment analysis on user comment. In: Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA). 2017 Presented at: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA); Sep 12-14, 2017; Kuching, Malaysia. [doi: 10.1109/icsipa.2017.8120658]
81. Cunha AA, Costa MC, Pacheco MA. Sentiment analysis of YouTube video comments using deep neural networks. In: Artificial Intelligence and Soft Computing. Cham: Springer; 2019.
82. AbdelFattah M, Galal D, Hassan N, Elzanfaly D, Tallent G. A sentiment analysis tool for determining the promotional success of fashion images on Instagram. Int J Interact Mob Technol 2017 Apr 11;11(2):66. [doi: 10.3991/ijim.v11i2.6563]
83. Ortigosa A, Martín JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. Comput Human Behav 2014 Feb;31:527-541. [doi: 10.1016/j.chb.2013.05.024]
84. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist J. Understanding the demographics of twitter users. In: Proceedings of the Fifth International Conference on Weblogs and Social Media. 2011 Presented at: Fifth International Conference on Weblogs and Social Media; Jul 17-21, 2011; Barcelona, Catalonia, Spain URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14168

Abbreviations

AR: augmented reality
HCW: health care worker
IoT: Internet of Things
MR: mixed reality
PPE: personal protective equipment
SVM: support vector machine
TF-IDF: term frequency–inverse document frequency
VR: virtual reality

XSL•FO
RenderX

Edited by N Zary; submitted 27.01.22; peer-reviewed by Z Shokri Varniab, R Gore, J Li, E Alam; comments to author 13.05.22; revised version received 27.05.22; accepted 12.06.22; published 04.08.22.

Please cite as:
Jeong H, Bayro A, Umesh SP, Mamgain K, Lee M
Social Media Users' Perceptions of a Wearable Mixed Reality Headset During the COVID-19 Pandemic: Aspect-Based Sentiment Analysis
JMIR Serious Games 2022;10(3):e36850
URL: https://games.jmir.org/2022/3/e36850
doi:10.2196/36850
PMID:35708916

©Heejin Jeong, Allison Bayro, Sai Patipati Umesh, Kaushal Mamgain, Moontae Lee. Originally published in JMIR Serious Games (https://games.jmir.org), 04.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Breathing as an Input Modality in a Gameful Breathing Training App (Breeze 2): Development and Evaluation Study

Yanick Xavier Lukic[1], MSc, MA; Gisbert Wilhelm Teepe[1], MSc; Elgar Fleisch[1,2], PhD; Tobias Kowatsch[1,3,4], PhD

[1]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Institute of Technology Management, University of St.Gallen, St.Gallen, Switzerland

[3]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[4]School of Medicine, University of St.Gallen, St.Gallen, Switzerland

Corresponding Author:
Yanick Xavier Lukic, MSc, MA
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich,
Switzerland
Phone: 41 446328638
Email: ylukic@ethz.ch

Abstract

Background: Slow-paced breathing training can have positive effects on physiological and psychological well-being. Unfortunately, use statistics indicate that adherence to breathing training apps is low. Recent work suggests that gameful breathing training may help overcome this challenge.

Objective: This study aimed to introduce and evaluate the gameful breathing training app Breeze 2 and its novel real-time breathing detection algorithm that enables the interactive components of the app.

Methods: We developed the breathing detection algorithm by using deep transfer learning to detect inhalation, exhalation, and nonbreathing sounds (including silence). An additional heuristic prolongs detected exhalations to stabilize the algorithm's predictions. We evaluated Breeze 2 with 30 participants (women: n=14, 47%; age: mean 29.77, SD 7.33 years). Participants performed breathing training with Breeze 2 in 2 sessions with and without headphones. They answered questions regarding user engagement (User Engagement Scale Short Form [UES-SF]), perceived effectiveness (PE), perceived relaxation effectiveness, and perceived breathing detection accuracy. We used Wilcoxon signed-rank tests to compare the UES-SF, PE, and perceived relaxation effectiveness scores with neutral scores. Furthermore, we correlated perceived breathing detection accuracy with actual multi-class balanced accuracy to determine whether participants could perceive the actual breathing detection performance. We also conducted a repeated-measure ANOVA to investigate breathing detection differences in balanced accuracy with and without the heuristic and when classifying data captured from headphones and smartphone microphones. The analysis controlled for potential between-subject effects of the participants' sex.

Results: Our results show scores that were significantly higher than neutral scores for the UES-SF (W=459; P<.001), PE (W=465; P<.001), and perceived relaxation effectiveness (W=358; P<.001). Perceived breathing detection accuracy correlated significantly with the actual multi-class balanced accuracy (r=0.51; P<.001). Furthermore, we found that the heuristic significantly improved the breathing detection balanced accuracy ($F_{1,25}$=6.23; P=.02) and that detection performed better on data captured from smartphone microphones than than on data from headphones ($F_{1,25}$=17.61; P<.001). We did not observe any significant between-subject effects of sex. Breathing detection without the heuristic reached a multi-class balanced accuracy of 74% on the collected audio recordings.

Conclusions: Most participants (28/30, 93%) perceived Breeze 2 as engaging and effective. Furthermore, breathing detection worked well for most participants, as indicated by the perceived detection accuracy and actual detection accuracy. In future work, we aim to use the collected breathing sounds to improve breathing detection with regard to its stability and performance. We also plan to use Breeze 2 as an intervention tool in various studies targeting the prevention and management of noncommunicable diseases.

XSL•FO
RenderX

(JMIR Serious Games 2022;10(3):e39186) doi:10.2196/39186

KEYWORDS

breathing training; serious game; biofeedback; digital health; mobile health; mHealth; mobile phone; machine learning; deep learning; transfer learning; neural networks

Introduction

Background

Noncommunicable diseases (NCDs) are a substantial global health and economic burden [1-3]. Slow-paced breathing training is positively associated with physiological [4-6] and psychological [7-9] well-being. Thus, breathing training can play a role in interventions targeting NCDs. For example, slow-paced breathing training may induce relaxation and help counteract stress [8]. It can also improve cardiac functioning [10], potentially enabling improved treatment of cardiovascular diseases, the leading cause of death worldwide [1]. Furthermore, it can strengthen respiratory muscles, rendering it relevant for the treatment of respiratory diseases such as asthma [4] and chronic obstructive pulmonary disease [11].

Slow-paced breathing training generally aims at guiding people to breathe with 5.5 to 6 breaths per minute (BPM) [6]. People may be able to maximize their personal effects by adjusting the BPM. For example, an untrained person may achieve better results by breathing with >6 BPM, whereas a well-trained person may want to breathe with <6 BPM. Nevertheless, 6 BPM are generally used as this appears to work well for most people, which results in 1 complete breathing cycle every 10 seconds. A breathing cycle consists of an inhalation, an exhalation, and up to 2 pauses in between. The duration of these individual phases is an area of active research. Investigating these separate phases is relevant as inhalation is associated with the sympathetic nervous system by inhibiting vagal outflow, and exhalation is associated with the parasympathetic nervous system by restoring vagal outflow [12,13].

Consequently, breathing patterns typically use equal inhalation and exhalation durations to balance the sympathetic and parasympathetic activity or prolonged exhalations to emphasize the parasympathetic nervous system. Although both approaches effectively induce relaxation [14], related work argues that a prolonged exhalation achieves more substantial relaxation effects. In contrast, other related work has found equal durations of inhalation and exhalation phases best suited to attain psychophysiological coherence [12,13,15].

The positive effects of breathing training have sparked the development of various breathing guidance apps [16]. Nevertheless, although these apps receive much attention and are downloaded by many users, the use statistics show that adherence is low [16,17]. The problem of nonadherence and lacking engagement is present in various domains, and different works hypothesize gamification as a potential solution [18,19]. In addition, for breathing training, various mobile [20-22], desktop [23], and virtual reality [24,25] applications that use some gameful elements have been conceptualized and developed. However, many apps are not interactive and do not include biofeedback during training, which is surprising as research indicates increased effectiveness of breathing training when biofeedback is used [26-29].

Nevertheless, breathing training apps exist that include biofeedback mechanisms based on heart rate variability (HRV) [30] or breathing [21-23,31]. Although HRV-based biofeedback offers feedback on the biosignal often targeted by breathing training [32], it only provides deferred feedback. It is also challenging to measure HRV without additional hardware. Thus, HRV biofeedback is neither scalable nor well suited as sole input for a gameful experience that requires feedback loops with latencies of <1 second.

By contrast, breathing-based biofeedback can be instantaneous and is the direct signal that the guidance in breathing training apps tries to change to then affect HRV. However, approaches using breathing-based biofeedback are so far limited to breathing training in controlled environments [21] and early prototypes [22]. One of the first apps to go in this direction was Breeze (Centre for Digital Health Interventions) [21]. The first version of Breeze featured a single environment where users accelerate a sailboat by correctly following a fixed breathing pattern. The effectiveness of Breeze in increasing HRV in individuals was shown in the laboratory [33].

Furthermore, Lukic et al [33] evaluated the effect of Breeze's visualization and visual breathing training guidance on participants' intrinsic experiential value. The results showed a significant increase in intrinsic experiential value when the gameful visualization was used compared with a standard guidance visualization while maintaining the same perceived effectiveness (PE) [34]. Nevertheless, investigations showed that the breathing phase detector used, enabling interactivity in Breeze, was very prone to noise and differences in individuals' breathing sounds and was overfitted on the data set used [21]. Research by Islam et al [35] extended the idea of breathing phase detection from breathing training–specific breathing to regular breathing. They focused on monitoring and diagnosis as measuring breathing phases in normal breathing has been motivated for diagnostic purposes [36,37]. Generally, monitoring and diagnosis are popular areas of research regarding breathing detection. Although Islam et al [35] aimed to monitor breathing phases during rest, others tried to detect breathing rates during sleep [38] and physical exercise [39] through smartphone and headphone microphones. However, as breathing training guides breathing, it does not make sense to investigate breathing phase durations and breathing rates during this time with diagnostic intentions. Breathing detection in the context of breathing training aims to provide interactive feedback to users to increase engagement and give them a tangible assessment of their performance.

Objectives

This paper introduces Breeze 2, which has several new features, an improved appearance, and a novel breathing phase detection

XSL•FO
RenderX

algorithm. We designed Breeze 2 as a stand-alone training and an intervention component in multicomponent interventions. Breeze 2 adds a slide-based tutorial to introduce users to breathing training and biofeedback mechanics. Furthermore, it allows for the setting of specific training parameters (ie, training duration and breathing pattern) according to the user's breathing training experience. It also adds procedural generation of the visual biofeedback environment so users always have a slightly different experience when performing breathing training. Consequently, this study has the following objectives: (1) to provide a detailed description of Breeze 2, a revised gameful breathing training app for smartphones; (2) to introduce and evaluate a novel real-time breathing phase detection approach based on deep transfer learning and an additional heuristic that prolongs detected exhalations to stabilize the algorithm's predictions; and (3) to evaluate perceived engagement, PE, and breathing detection performance in a laboratory setting with 30 participants.

Methods

Design and Implementation

Overview

Breeze 2 provides visual breathing guidance through gameful visualizations. Furthermore, it incorporates interactive components that aim to make the training more engaging and provide valuable feedback to users on their breathing training performance. We outline the details of the revised user interface design and breathing detection algorithm of Breeze 2 in the following sections.

Concept and Design

Overview

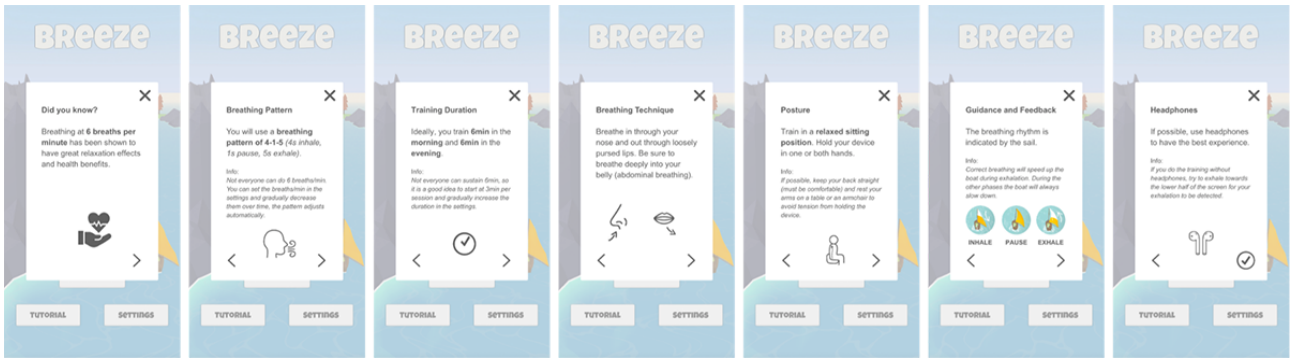
In contrast to many other breathing training systems, Breeze 2 does not rely on abstract shapes but uses a tangible setting that allows for the design of the interactive components in a relatable way. A sailboat that continuously moves forward slowly guides the breathing training. Depending on the user's ability to match the guiding breathing pattern, the exhalation triggers stronger winds in the experience, and the sailboat accelerates. This way, the sailboat travels a larger distance over the duration of the training, which allows for the use of the traveled distance as a condensed measure of training performance aside from more precise measurements such as the timing accuracies on exhalations.

In addition to the breathing training itself, Breeze 2 also offers a tutorial, adjustable training parameters, and procedural generation to vary the shown environment for every breathing training session.

Tutorial

Breeze 2 uses a simple slide-based tutorial (Figure 1) that quickly introduces the user to the benefits of slow-paced breathing training and briefly explains the guidance and interaction components. When a user opens Breeze 2 for the first time, the start button on the home screen is disabled. Once the user has completed the tutorial, the start button is enabled, and the user can start a training session.

Figure 1. Slide-based tutorial as implemented in Breeze 2. It provides high-level information on the benefits of slow-paced breathing training and its biofeedback mechanics.

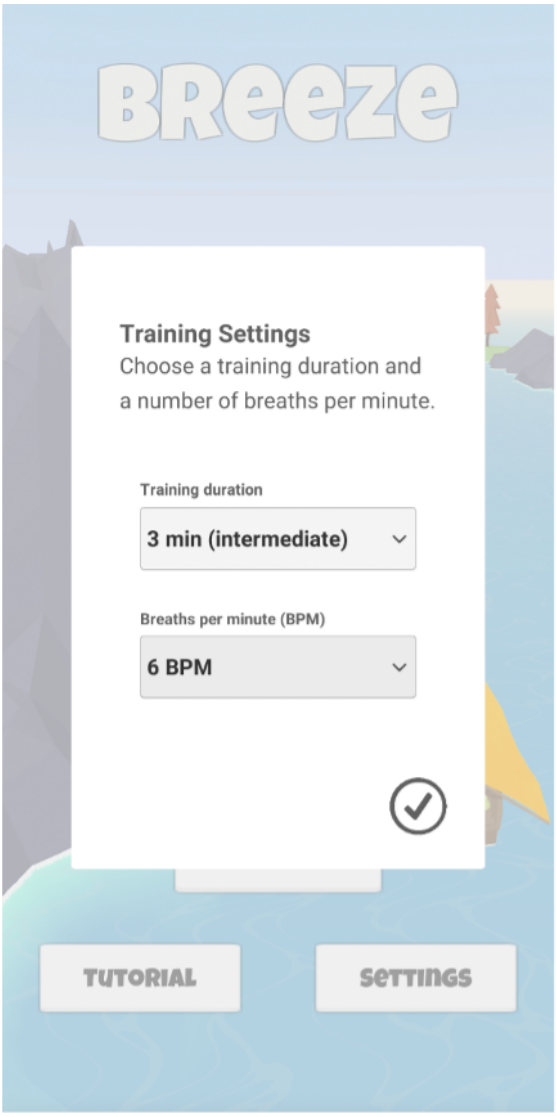


Training Parameters

Breathing training mainly consists of 2 parameters: the training duration and the breathing pattern. Users can adapt both according to their preferences. Breeze 2 supports this process by labeling the possible durations according to their required level of breathing training experience. We chose the breathing training parameters based on feedback from medical professionals working with biofeedback-guided breathing exercises as patient treatment. Regarding training durations, the user can choose between *2 (beginner)*, *3 (intermediate)*, and *5 (expert)* minutes. The breathing patterns in breathing training usually take the form of *inhalation-pause-exhalation-pause*. Breeze 2 uses the pattern *4-1-5-0* and adjusts the inhalation and exhalation duration linearly to match the desired BPM. For example, if 6 BPM is selected, the breathing pattern follows 4, 1, 5, and 0 seconds of inhalation; first pause; exhalation; and second pause. However, if 7 BPM is selected, the pattern follows 3.37, 1, 4.21, and 0 for the 4 phases. As a standard selection, we used 6 BPM. Figure 2 illustrates Breeze's settings screen.

XSL•FO
RenderX

Figure 2. Settings screen where the users can set the training duration and the breaths per minute according to their preferences.



Voice Commands

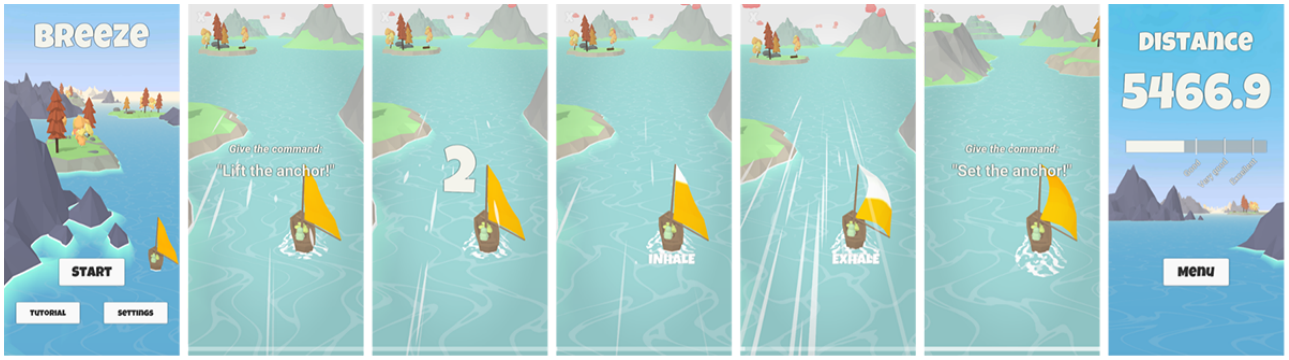
Breeze 2 allows for the enabling of voice commands to start and end breathing training. We implemented this feature to enable future studies using Breeze to gather and analyze voice features regarding the studies' outcomes. If a user speaks for a specified time, Breeze 2 approves the command. We can configure the content and expected durations of commands according to the studies' needs. Consequently, Breeze 2 does not check whether the user said the correct words but that they said something. This way, it is less error-prone and allows users to speak more naturally, with the caveat that it is required to trust the user to speak the displayed words. We used a pretrained Yet Another Mobile Network (YAMNet) model for the necessary voice detection [40].

Interaction During Training

Users can initiate a breathing training session from the start menu. When the training starts, the view changes to the training mode. Initially, users see a sailboat floating on a river from behind. The sailboat first stands still. For the next step, the users need to read 3 voice commands aloud, after which a countdown starts. At the end of the countdown, the guidance breathing pattern starts. The BPM parameter the user sets determines how long the individual phases are in seconds. An animation on the sail of the sailboat represents the separate phases. During the first 5 breathing cycles, Breeze 2 also indicates the breathing phases through an additional text label below the sailboat. In the beginning, the sailboat moves forward at a slow constant speed. The users must then adapt their breathing to the breathing pattern by following the guidance system. The more accurately the users follow the pattern, the more strongly the sailboat accelerates during the exhalation phase. A correctly timed exhalation triggers a wind animation that propels the sailboat forward. During the inhalation and pause phases, the sailboat's speed slowly decreases until it reaches the constant base speed. Once users complete a training session, they again speak 3 voice commands aloud. After that, the users see a new screen showing the reached distance and the breathing accuracy over the completed session. Figure 3 depicts a complete training session. A screen recording of a complete session without voice commands can be found in Multimedia Appendix 1.

XSL•FO
RenderX

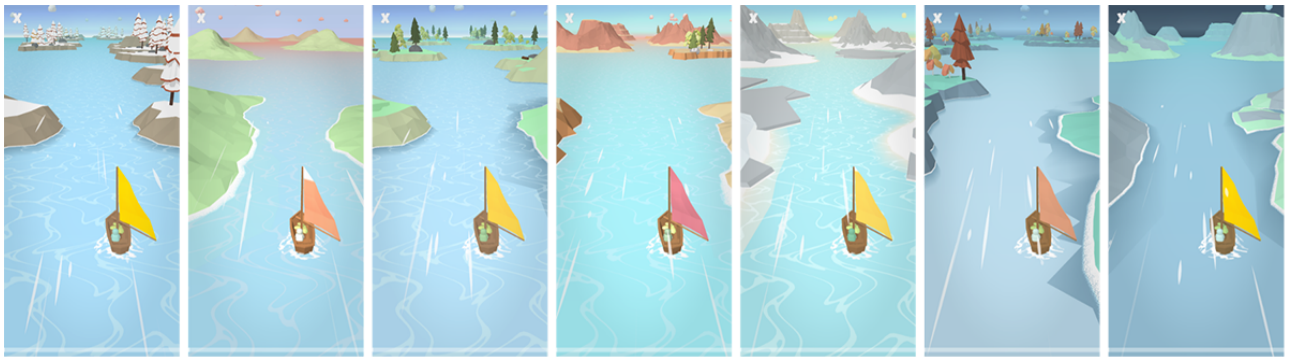
Figure 3. An entire session of Breeze 2 (from left to right): start screen, starting voice commands, countdown, inhalation phase, exhalation phase, ending voice commands, and final screen.



Procedural Generation

As users should perform breathing training regularly, it is essential to provide visual variety every time to support long-term adherence. Breeze 2 uses simple procedural generation that varies the environment's appearance and lighting during training sessions to achieve this visual variety. The procedural generation works with predefined configurations that include groups of 3D models and different coloring and lighting schemes. We handcrafted these configurations to ensure that they fit together. Every session, the app randomly chooses one of the configurations. Subsequently, the procedural generation places island models assigned to this configuration along the travel path of the sailboat at random locations. Furthermore, this configuration's coloring and lighting scheme are chosen and applied to the scene. Figure 4 illustrates such generated scenes, including islands, coloring schemes, and landscapes.

Figure 4. A selection of procedurally generated landscapes during breathing training sessions.



Background Sounds

If the users use Breeze 2 without headphones, it does not have any sound to not interfere with breathing detection. Otherwise, Breeze 2 plays a peaceful background sound during the training session. It combines soft water sounds with subtle animal sounds such as birds. The background sound is audible in the screen recording in Multimedia Appendix 1.

Stand-alone and Intervention Component

Breeze 2 can be used either as a stand-alone breathing intervention or as an intervention component for multicomponent interventions. For the former, a start screen allows the users to set training parameters via the settings menu (Figure 2) and a simple slide-based tutorial (Figure 1). When built as an intervention component for a multicomponent intervention, training parameters can also be handed over as parameters to Breeze 2, and the training may start right away. The handing over of parameters is useful if the multicomponent intervention (eg, a smartphone-based holistic lifestyle intervention) already features a tutorial and the possibility to choose training parameters (eg, via chatbot).

Implementation

We used the Unity real-time development platform (version *2020.3.4f1*; Unity Technologies) to implement Breeze 2. All 3D models were custom creations or acquired through the Unity Asset Store. For 3D model creation and modification, we used the 3D modeling software Blender (Blender Foundation). The background sound was downloaded from Freesound [41] and was available under the Creative Commons Zero license.

Real-time Breathing Detection

Overview

The aim of breathing detection for interactive breathing training is to detect inhalation and exhalation phases as fast as possible to enable real-time feedback. Consequently, the detection algorithm must distinguish these 2 phases and all nonbreathing sounds. Previous work that tried to detect breathing phases during breathing training [21] and natural breathing at rest [35] used preceding breathing detection gates that check inputs for breathing sounds before passing them to the model that classifies only breathing-related classes. Shih et al [21] tried to detect breathing in close to real time and used a breathing gate that works on 1-second clips. Aside from inhalations and exhalations,

XSL•FO
RenderX

they also tried to detect breathing pauses. However, as they applied a sequence model and wanted to account for clips that included different phases, they split pause phases into inhalation-pause and exhalation-pause.

Islam et al [35] focused on breathing monitoring and diagnosis and, thus, used a 1-minute breathing gate. They also made the simplifying assumption that breathing is continuous and has no pauses. This focus and assumption allowed them to reduce the problem to a 2-class problem for their primary model with the classes inhalation and exhalation.

We did not apply a sequence model and aimed for real-time predictions. Thus, we could not work with input durations of 1 minute. Furthermore, we argue that a single model approach can be beneficial as the primary model then does not only come into contact with a limited domain. Consequently, we used only 1 model and defined the problem as a 3-class problem with the classes inhalation, exhalation, and nonbreathing sounds (including silence).

Similar to Shih et al [21], this work focuses on applying a breathing detector in breathing training guided by an app running on a smartphone. This comes with a caveat as, when detecting exhalations using a smartphone's microphone, it is essential to distinguish between detecting the exhalation from sound alone and the airflow itself. Users may exhale toward the device during training, leading to disturbances in the audio recording usually produced by wind. Identifying these disturbances is especially relevant if slow-paced breathing is combined with pursed-lip breathing as the air stream is becoming more focused this way. Therefore, the model should still detect the resulting disturbance sounds as exhalations resulting in 2 subtypes of the exhalation class, which we call acoustic and airflow exhalations in this paper. However, the model should assign samples from both subtypes to the exhalation class regardless of whether they are acoustic or airflow exhalations.

Data Set

We formed the data set used for training, validation, and preliminary testing from 3 separate data sets. The first consists of acoustic breathing sounds, the second consists of exhalation-generated airflow disturbance sounds, and the third consistis of environmental sounds.

We used a subset of the data set from Shih et al [21] for acoustic breathing sounds. We only used the recordings produced by the RØDE NT1000 microphone, which had the best quality. Furthermore, we manually selected only recordings that contained audible breathing and little or no constant background noise, which resulted in audio data from 20 participants. As breathing training is often performed by inhaling through the nose and exhaling through the mouth, we only included these sounds for the breathing data set. Data from the first 80% (16/20) of the participants served as training and validation data. The validation set contained the last 3 breathing cycles by a participant, whereas we used the remaining breathing cycles for training. Data from the remaining 20% (4/20) of the participants served as test data that we used to ensure model testing on only data from unseen individuals.

As the data set from Shih et al [21] only contains acoustic breathing sounds, we recorded new data for exhalations that produce disturbances in the recording through airflow. Given that these disturbance sounds are the same as those produced by wind hitting the microphone's membrane, they are mostly independent of the individual's breathing sound. The smartphone used has a more significant influence as the microphone's position and the device's overall architecture influence how much air reaches the microphone's membrane. Consequently, a male and a female participant performed three 2-minute breathing training sessions. The 2 participants used different smartphones without headphones for the training sessions. Both participants exhaled toward the device during training. The exhalation sounds were then manually extracted from the resulting recordings. To ensure that the airflow sounds were independent of the individual, we only included the samples produced by the male participant in the training and validation sets and used the samples from the female participant in the test set.

For nonbreathing sounds, we used the data set ESC-50 [42], which entails 50 classes of environmental sounds. Every recording is 5 seconds long, with 40 recordings per class. We excluded all breathing sounds from the data set and used folds 1, 2, and 3 for the training, validation, and test set, respectively. We also used nonbreathing sounds and silence from the breathing sound data set from Shih et al [21]. They were distributed in the same way as the breathing sounds in the training, validation, and test sets. We used these nonbreathing sounds and silence to ensure the model did not use the environmental characteristics of the recordings to distinguish between breathing and nonbreathing sounds.

All recordings in the data set were then cut into 0.195-second–long nonoverlapping clips. Table 1 describes the resulting composition of the data set.

Table 1. Data set composition used for training, validation, and the testing of the developed model.

Class	Samples, n		
	Training	Validation	Testing
Exhalation (acoustic)	4574	941	753
Exhalation (airflow)	418	82	455
Inhalation	2470	478	663
Nonbreathing (ESC-50 data set)	9800	9800	9800
Nonbreathing (laboratory)	1952	1952	552

XSL•FO
RenderX

Transfer Learning Approach

Overview

We used a pretrained YAMNet [40] model as the basis for transfer learning. YAMNet is a convolutional neural network based on the MobileNetV1 [43] architecture trained on the AudioSet data set [44] to classify 521 classes. Transfer learning refers to using a pretrained model or relevant parts of it and fine-tuning it on a related problem [45].

Preprocessing

The audio samples were preprocessed to fit the YAMNet requirements. Specifically, we resampled the audio to 16 kHz mono. Here, we introduced a step specific to our problem. YAMNet uses a minimum of 15,600 data points as input, which corresponds to 0.975 seconds (internally, it works with 0.96-second patches but requires additional samples to compute the final Short-time Fourier transform window [40]). However, it is questionable whether 1 second is fast enough for real-time feedback that should be perceived as immediate. Research in touch-based systems indicates that commercial touch screens yield latencies of up to 200 ms [46] and that perceivable latency lies between 2 and 100 ms [47]. To the best of our knowledge, no such research exists for breathing inputs. We hypothesize that the perceived latency in breathing-based systems is not as sensitive as in touch-based systems.

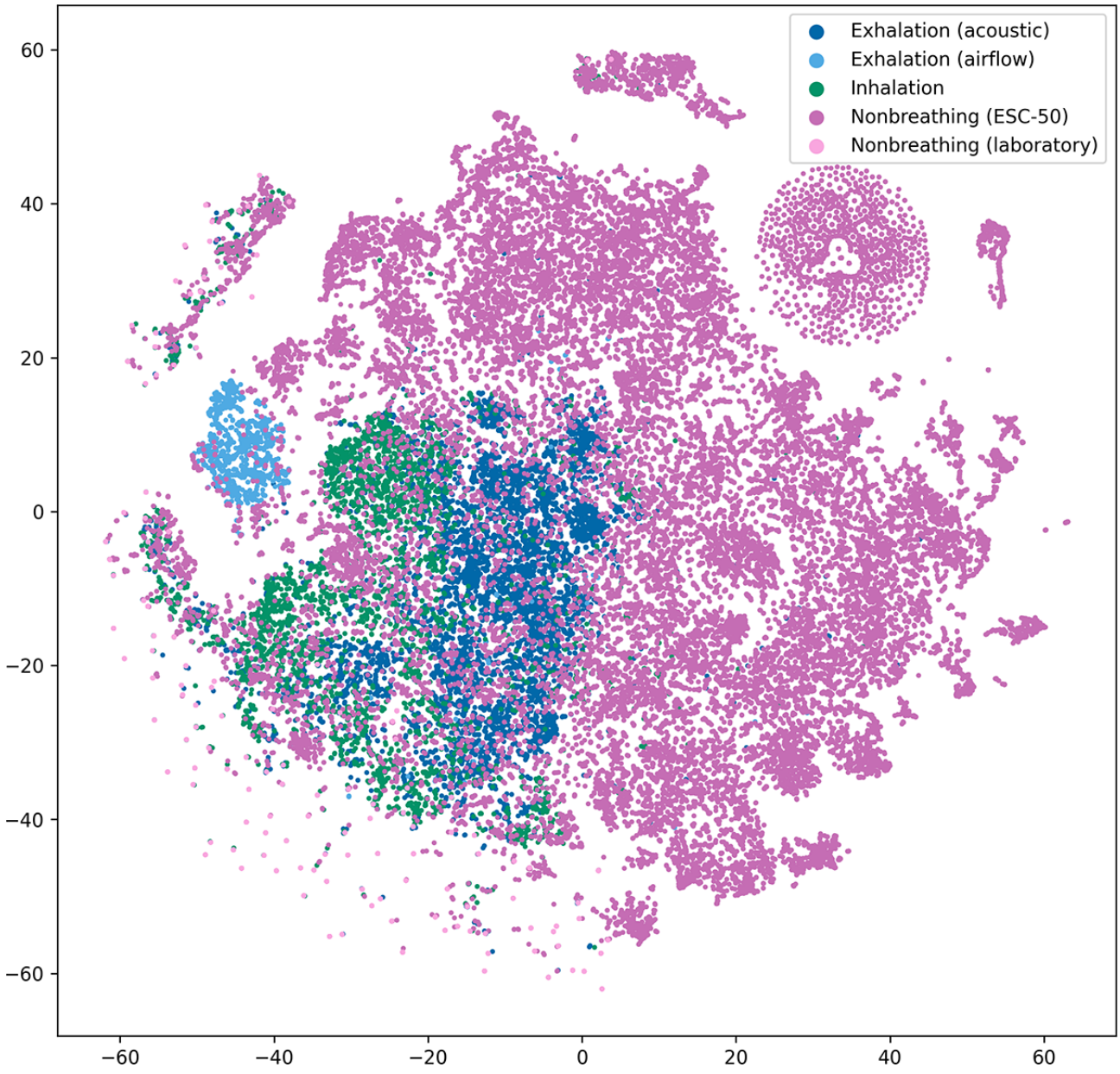
Consequently, we aimed for an input size of >100 ms but still significantly <1 second to ensure that the input contained enough information but could still provide feedback that users may perceive as immediate. We decided to use 0.195 seconds as input size, corresponding to 3120 samples and one-fifth of a YAMNet input. We then concatenated this snippet to arrive at the total input for YAMNet. Not just padding the signal with some constant value ensures that inputs containing distinct sound sources are as different as possible from, for example, quiet environments. We then calculated a mel spectrogram with a window and hop size of 25 and 10 ms, respectively. The mel spectrogram consisted of 64 mel bins covering the range of 125 to 7500 Hz. Finally, we calculated the log mel spectrogram by calculating $\log(S+0.001)$, where S is the mel spectrogram.

Feasibility Check

To assess whether the embeddings of YAMNet captured features that allowed for distinguishing between inhalation, exhalation, and nonbreathing sounds, we used the *t*-distributed stochastic neighbor embedding method [48]. We calculated embeddings for all samples in the data set, resulting in 1024-dimensional embeddings that we then reduced to 2D embeddings using *t*-distributed stochastic neighbor embedding (with Euclidean metric). We then visualized the 2D embeddings in a scatter plot and manually inspected these representations (Figure 5). We observed that the airflow exhalation samples clustered separately from the acoustic breathing sounds. In addition, airflow exhalations clustered together regardless of person and device. The acoustic exhalations also clustered together but partially overlapped with the inhalations. The visualization also showed that nonbreathing sounds formed various clusters among themselves as the ESC-50 data set contained different types of sounds. The samples from the laboratory containing nonbreathing sounds were also spread across a wide range but separated quite clearly from breathing sounds while partly overlapping with various ESC-50 clusters.

XSL•FO
RenderX

Figure 5. Visualized Yet Another Mobile Network (YAMNet) embeddings for the complete data set. We used t-distributed stochastic neighbor embedding to reduce the high dimension of the embeddings. ESC-50: Dataset for Environmental Sound Classification.



Training

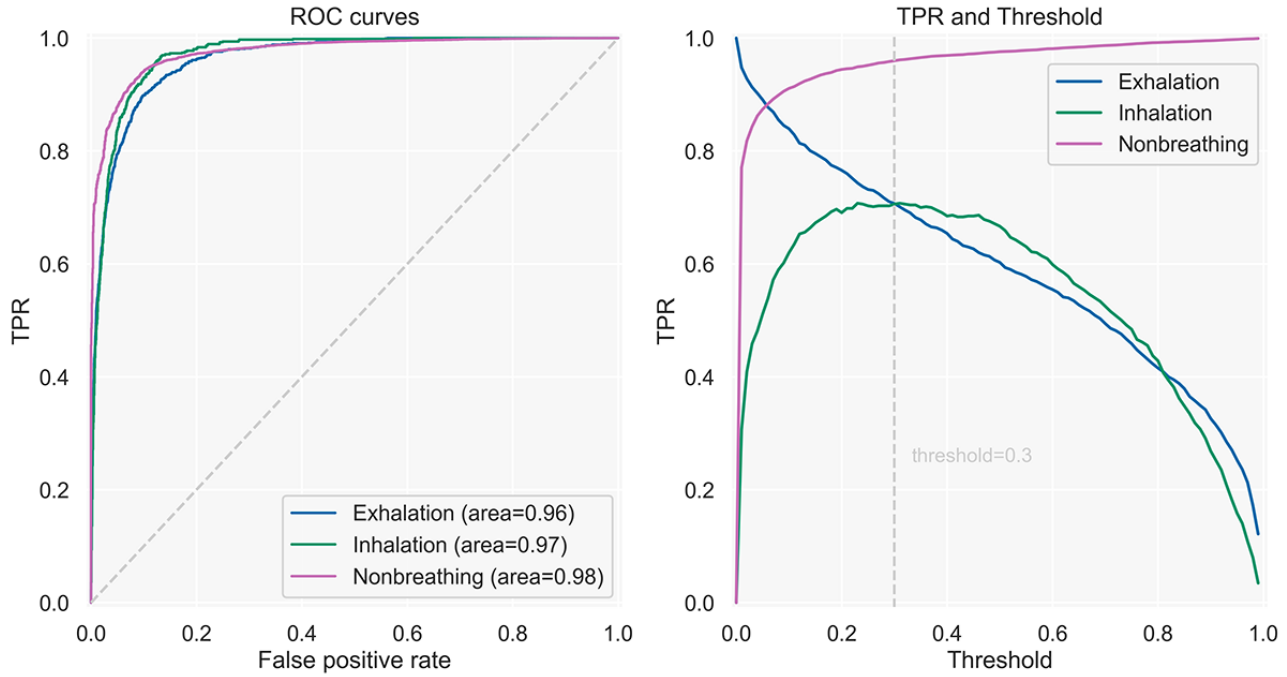
For transfer learning, we used the 1024-dimensional embeddings generated by YAMNet and fed them into a small neural network consisting of 2 fully connected layers with 32 and 3 units. The first layer applied the swish [49] activation function, and the output layer applied the softmax function. We trained the algorithm using mini-batch gradient descent with the Adam optimizer and categorical cross-entropy as loss function. Mini-batch size was set to 32. Our manual testing showed that the algorithm usually started to overfit on the training set after 5 to 10 epochs. We then used early stopping with patience of 10 epochs and restored the best weights according to the lowest loss reached on the validation set. Even though the used data set was strongly imbalanced, we did not use any balancing approaches as there is more diversity in nonbreathing sounds than in exhalations and inhalations. This way, we wanted to discourage false positives on breathing sounds. Otherwise, the detector may yield problems in not perfectly quiet environments. Using this transfer learning approach, we created 1000 models and chose the 3 that reached the lowest loss on the validation set to be combined as an ensemble. The ensemble applied soft voting with equal model weights (the class with the maximum sum of probabilities is chosen). This was done to slightly increase the performance and stability of the model's predictions.

Evaluation

To evaluate the model, we used the unseen test set. We investigated the receiver operating characteristic (ROC) curves; confusion matrix; and the precision, recall/sensitivity, specificity, F_1 score, and balanced accuracy metrics. The ROC curves (Figure 6) yielded areas under the curve of 0.96, 0.97, and 0.98 for exhalation, inhalation, and nonbreathing sounds, respectively, indicating good discrimination capacity between all classes.

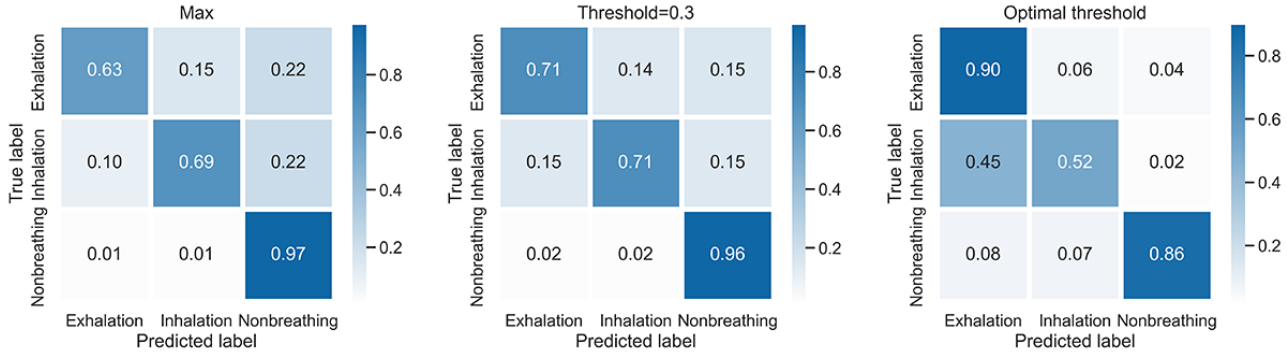
XSL•FO
RenderX

Figure 6. The left diagram depicts the model's ROC curves for exhalation, inhalation, and nonbreathing sounds on the test set. The right chart shows the TPR for all classes when applying different thresholds for the breathing classes. We used the same threshold for inhalations and exhalations but applied it first to exhalations. We used the visual inspection of the right chart to determine a suitable threshold for the breathing classes. ROC: receiver operating characteristic; TPR: true positive rate.



To identify thresholds for exhalation and inhalation detection, we applied 2 approaches. First, we calculated the optimal thresholds for inhalation and exhalation individually by selecting the threshold that yielded the highest Youden *J* statistic [50]. Second, we plotted the true positive rates for several thresholds and established an appropriate threshold via visual inspection that yielded a balance between the 3 classes (Figure 6). We applied the threshold first for exhalation and then for inhalation and, if they did not apply, the model yielded nonbreathing. We found the threshold of 0.3 to strike a reasonable balance between the 3 classes. Figure 7 shows the confusion matrices for the standard threshold (maximum probability), the optimal thresholds, and the threshold of 0.3 for the test. We concluded that the threshold of 0.3 reached a better balance and, thus, discrimination among classes.

Figure 7. Confusion matrices showing the model results on the test data set applying 3 different thresholds for the breathing classes. From left to right: max (the class with the highest prediction probability is selected), threshold=0.3 (the threshold of 0.3 is applied first to exhalation and then to inhalation), and optimal threshold (different optimal thresholds for exhalation and inhalation are applied, as determined by the Youden J statistic).



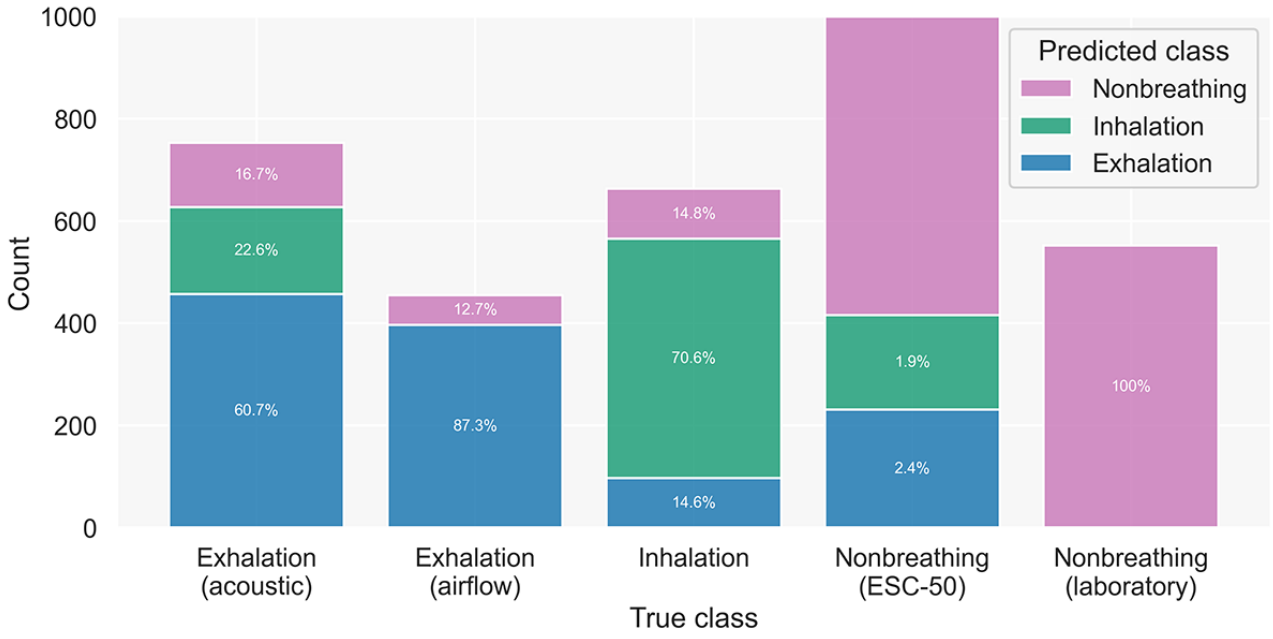
Consequently, we chose this threshold for further evaluation. The precision, recall/sensitivity, specificity F_1 score, and balanced accuracy metrics for this model are provided in Table 2. They show that the model best detects nonbreathing sounds, the most dominant class in the training set. The confusion matrices also show that the model more often misclassified exhalation and inhalation samples as nonbreathing sounds than as the wrong breathing phase. To gain further insights regarding the correct and incorrect classifications, we visualized test set classifications for all subclasses of the 3 main classes (Figure 8). The figure indicates that acoustic exhalations are similarly often misclassified as inhalations and nonbreathing sounds, whereas airflow exhalations are only misclassified as nonbreathing sounds. Inhalations yield a similar result as acoustic exhalations. For nonbreathing sounds, the samples from the laboratory appear to be easily distinguishable by the model. Sound samples from ESC-50 yield some misclassifications, with most being exhalations. Nevertheless, the misclassifications for nonbreathing sounds are only a small portion of all samples of this class.

XSL•FO
RenderX

Table 2. Performance metrics of the model using a threshold of 0.3 for the breathing classes on the data from the test set.

Class	Precision	Recall/sensitivity	Specificity	F_1 score	Balanced accuracy
Exhalation	0.72	0.71	0.97	0.71	0.84
Inhalation	0.57	0.71	0.97	0.63	0.84
Nonbreathing	0.97	0.96	0.85	0.97	0.90
Average	0.75	0.79	0.93	0.77	0.86

Figure 8. Histogram showing test set classifications by the model using a 0.3 threshold for breathing sounds. We split the data according to the subsets. As the ESC-50 subset of the nonbreathing sounds is substantially larger (9800 samples) than the other subsets, we cut off the diagram at 1000 samples. ESC-50: Dataset for Environmental Sound Classification.



Model Inference Time Measurement

We conducted a basic performance measurement of the resulting model on 3 smartphones. The main objective was to verify that the model could perform inference in <0.195 seconds, which corresponds to the input duration of the audio signal. As it can be expected that more powerful devices allow for faster inference, we focused on low- to midrange Android smartphones from different device manufacturers. We used the TensorFlow (Google Brain Team) Android benchmark app [51] to measure the performance of our model after conversion to a TensorFlow Lite model. The benchmark app simulates the model's execution within an actual Android app. Thereby, it is ensured that Android's scheduler treats the thread and process priorities of the model inference like those of a foreground app. We ran inference time measurements on a Samsung Galaxy S10e, a OnePlus 6, and a Huawei P30 Lite. All devices were factory reset before the benchmark app was installed. Furthermore, auto-lock was disabled to ensure that the devices did not switch to low-power mode during measurements. No hardware acceleration was used (ie, the use of a graphics processing unit, the NNAPI, the XNNPACK, and Hexagon was disabled in the benchmark app). We performed the measurements for 1, 2, and 4 central processing unit threads. For every device and thread configuration, we ran 100 inferences with 1 warm-up run. As the model was continuously running in the target use case, the warm-up and initialization times were neglectable and, thus, not reported. Table 3 lists the average measurements and their SDs. On the Samsung Galaxy S10e, measurements became unstable when using >2 central processing unit threads.

Table 3. Inference timings for the model on a small battery of Android smartphones.

Device	Inference time (μ; ms), mean (SD)		
	1 CPU[a] thread	2 CPU threads	4 CPU threads
Samsung Galaxy S10e	7.71 (0.10)	5.44 (0.09)	7.05 (1.26)
OnePlus 6	15.37 (0.02)	9.39 (0.06)	6.42 (0.10)
Huawei P30 Lite	24.17 (0.13)	15.19 (0.99)	11.61 (1.32)

[a]CPU: central processing unit.

XSL•FO
RenderX

Consequently, we used 2 threads for model inferences in Breeze 2. The measurements showed that the model can make an inference on all tested smartphones below the input size of 1 sample (0.195 seconds). Therefore, it should be able to monitor an incoming audio stream seamlessly.

Implementation in Breeze 2

As the feedback mechanism focuses mainly on the exhalation, we used an additional heuristic for exhalation detection. For this heuristic, we exploited the high precision of exhalation detection and the fact that detection runs every frame (approximately 30 times per second). If an exhalation was detected, we used a delay of 300 ms, during which the heuristic set the detected phase to exhalation. Every time the model detected an exhalation, the heuristic reset the delay. This way, once an exhalation was detected, the user could receive an immediate reaction, which the heuristic maintained for at least 300 ms. We used this maintaining of detected exhalations to counter the expected moderate detection performance on exhalations when deploying the model in settings with higher background noise levels. Furthermore, it prevented too abrupt changes between sailboat acceleration and deceleration. Consequently, the heuristic was specific to the feedback loop used of the sailboat accelerating during correct exhalations.

User Study

We aimed to evaluate this new iteration of Breeze and evaluate the breathing detection algorithm on new and realistic data through a laboratory study.

Participants

We recruited 30 participants (women: n=14, 47%; age: mean 29.77, SD 7.33 years). As our main interest was to collect breathing sounds and explore how certain aspects of Breeze 2 were perceived and how well breathing detection worked, the nature of the sample was not crucial. Consequently, we recruited participants mainly from ETH Zürich, but participation was open to all interested parties. However, participants had to be aged ≥18 years and not pregnant. Furthermore, they were required not to be taking any medication to treat depression, anxiety, or the main symptoms of mood disorders (such as low mood) and not to have any respiratory diseases such as asthma or chronic obstructive pulmonary disease. We aimed to balance female and male participants to account for potential differences in breathing sounds that may occur owing to physiological differences in respiratory function [52].

Materials

We measured *user engagement* using the User Engagement Scale Short From (UES-SF) [53]. This instrument consists of 4 subscales: *focused attention*, *perceived usability*, *esthetic appeal*, and *reward factor*. A total of 3 items measure each subscale.

The instrument to investigate *PE* of the breathing training consists of the following six items [54]: (1) *The breathing training facilitates relaxation*, (2) *The breathing training is pleasant to use*, (3) *It is easy to follow the breathing training instructions*, (4) *The breathing training effectively teaches how to breathe*, (5) *The breathing training is effective in reducing stress*, and (6) *The breathing training is effective in increasing attention to breath*. Each item was rated on a 5-point Likert scale (strongly disagree to strongly agree). To build the score for PE, we averaged the scores from all items. To construct the score for perceived relaxation effectiveness, we used the average of items 1 and 5.

Participants reported their *perceived breathing detection accuracy* with 2 independent items. The first one was adapted from the study by Efendic et al [55] and asked “How accurate is the breathing detection?” It was rated on a 7-point Likert scale (very inaccurate to very accurate). The second item asked “How much of your breathing did the breathing detection correctly detect?” Participants responded using a slider ranging from 0% to 100%. The questionnaires used in the study can be found in Multimedia Appendix 2.

We used 5 different smartphones in the study: *Samsung Galaxy S10e*, *OnePlus 6*, *Huawei P30 Lite*, *iPhone XR*, and *iPhone 11 Pro*. Each participant used only 1 smartphone, which was randomly assigned. All participants used Apple AirPods second generation [56] as headphones.

Procedure

After they signed the consent form at the start of the study, the participants received one of the smartphones with the stand-alone version of Breeze 2. The investigator then asked the participants to perform 2 breathing sessions with Breeze 2, one performed using headphones and the other without any additional hardware aside from the smartphone. Whether the participants started with or without headphones was randomly assigned. Each session was 3 minutes long. Before the first session, the investigator instructed the participants to read through the tutorial (Figure 1) and asked them to set the training duration and the breathing pattern to 3 minutes and 6 BPM, respectively. The investigator encouraged the participants to ask questions freely if the instructions provided in the app were not clear enough. We decided to allow such an additional explanation as assessing the quality of the tutorial was not a major objective of this study. After the first breathing session, the participants answered questions about their engagement (UES-SF) [53], the PE of the visualization [54], and the perceived accuracy of the breathing detection algorithm (adapted from the study by Efendic et al [55]) and provided their age and sex. Subsequently, they performed the second round of breathing training, after which they again answered the questions regarding the perceived accuracy of the breathing detection algorithm. If the participants wanted, they were allowed to interact with Breeze 2 for an additional 5 minutes, but this part was optional. Finally, the investigator encouraged participants to share feedback regarding Breeze 2 and the study. The sounds captured during breathing training were recorded for further offline analysis and future training data to refine the model.

Data Collection

Breeze 2 continuously monitors the breathing phase reference shown to the user during training and the breathing phases detected by the model used with and without the heuristic. This information is sampled every frame and, thus, usually results in 30 data points per second that are written to a log file.

XSL•FO
RenderX

However, this number fluctuated depending on the smartphone's computational power and the current scenery shown. Breeze 2 recorded and stored audio through the device's microphone during training sessions. When the participants used headphones because of the study design, Breeze 2 recorded the audio with the headphones' microphones. Breeze 2 tried to record with 44.1 kHz. However, operating system settings could overwrite this setting. In these cases, Breeze 2 recorded audio with at least 16 kHz, sufficient for the breathing detection model used. Breeze 2 similarly recorded the pre- and posttraining voice commands and yielded 2 separate recordings from the main breathing training recording. However, the voice commands were not further analyzed in this study.

After the data collection, 2 raters independently labeled the breathing training recordings as *exhalations* and *inhalations*. The raters did not manually label nonbreathing sounds. However, if a part of the recording was not assignable to an exhalation, an inhalation, or another sound, it was labeled as *unclear*. Unlabeled portions of the recording were then automatically labeled as nonbreathing sounds. A Cohen κ of 0.91 indicated near-perfect interrater reliability. Most mismatches came from slightly different label start and end times in the time-series data. Start or end time differences of >200 ms were manually inspected and merged, and others were merged by choosing the average of both raters. In case different class labels were assigned, either one of the raters' labels was chosen for the corresponding segment or it was marked as unclear. We then transformed the labeled data into a data set following the same steps as the training data. The resulting data set consisted of 20,753, 10,459, and 19,265 samples for exhalation, inhalation, and nonbreathing sounds, respectively.

Statistical Analyses

For the collected data, we formulated the following hypotheses: (1) the engagement score is higher than the neutral score (neither agree nor disagree; hypothesis 1); (2) the PE is higher than the neutral score (neither agree nor disagree; hypothesis 2.1); (3) the perceived relaxation effectiveness is higher than the neutral score (neither agree nor disagree; hypothesis 2.2); (4) the balanced detection accuracy of the model alone is lower than the balanced exhalation detection accuracy, including the heuristic (hypothesis 3.1); (5) the balanced detection accuracy is lower for sounds captured by headphone microphones than by smartphone microphones (hypothesis 3.2); (6) there is a difference in balanced detection accuracy for female and male participants (hypothesis 3.3); and (7) the perceived detection accuracy correlates with the actual balanced breathing detection accuracy (model including the heuristic; hypothesis 4).

To ensure construct reliability, we calculated the McDonald ω [57] for all subscales of the UES-SF and the overall user engagement score (UES), PE, and perceived relaxation effectiveness. For all subsequent hypothesis tests, we used an α level of .05. To test hypotheses 1, 2.1, and 2.2, we conducted Wilcoxon signed-rank tests against the neutral score of 3.0 on the UES-SF, the PE, and the perceived relaxation effectiveness. To account for the familywise error rate of PE and perceived relaxation effectiveness, we applied the Bonferroni correction to adjust the *P* values. To gain more insight, we conducted Wilcoxon signed-rank tests for the 4 subscales of the UES-SF and applied the Bonferroni correction to adjust the *P* values to account for the familywise error rate. For hypotheses 3.1 to 3.3, we calculated the balanced detection accuracy of the model, including the heuristic and the model alone based on data from the log files aggregated with the labels of the audio recordings. We used balanced accuracy as the heuristic should increase exhalation sensitivity while decreasing the specificity. It should also affect the sensitivity and specificity measure for the other 2 classes.

Consequently, we used multi-class balanced accuracy [58] as the dependent variable for this analysis as it includes all classes' specificity and sensitivity measures [59]. We then conducted a repeated-measure ANOVA with balanced accuracy as the dependent variable, the presence of the heuristic and the use of headphones as repeated-measure factors, and the participants' sex as a between-subject factor. The latter was included to account for any potential breathing sound differences between men and women because of physiological differences [52]. A Shapiro-Wilk test [60] verified the normal distribution of the data for all 4 cells: heuristic (W=0.93; P=.07), headphone (W=0.97; P=.68), heuristic and headphone (W=0.94; P=.14), and neither (W=0.97; P=.55). We tested the assumption of homogeneity of variances for all the sex-based subgroups within the cells using the Brown-Forsythe test, which is a more robust Levene test [61] using medians instead of means to calculate the center of each group [62]. The assumptions of homogeneity of variance were met for heuristic ($F_{1,25}$=1.75; P=.20), headphone ($F_{1,25}$=0.07; P=.79), heuristic and headphone ($F_{1,25}$=0.07; P=.80), and neither ($F_{1,25}$=0.79; P=.38). The assumption of sphericity was met as the repeated measures had only 2 levels. To investigate hypothesis 4, we conducted Pearson correlation tests between the actual balanced breathing detection accuracy and the perceived detection accuracy items separately. This tested whether the found correlations differed from 0. We then scaled the responses from the response values to be between 0 and 1 (divided by the maximum allowed value for each item) and plotted them with the balanced breathing detection accuracy in 2 Bland-Altman plots [63] to investigate the tendencies of the differences.

Model Evaluation

We also investigated the detection performance of the model (excluding the heuristic) offline on the collected audio recordings. This was done for 2 reasons; first, to obtain detailed insights into the model's detection performance without the heuristic used. Thereby, we obtained more information on the transferability of the model to potential other implementations where the heuristic would not be adequate. Second, this evaluation may serve as a baseline for future work as it was done in a standardized way offline on the collected breathing recordings. We considered the ROC curve of each class. Furthermore, we investigated the precision, recall/sensitivity, specificity, F_1 score, and balanced accuracy (average of sensitivity and specificity) metrics. We included all these metrics to provide a complete picture of the model's performance. Furthermore, we analyzed the detection performance for samples captured via smartphone and headphone microphones.

XSL•FO
RenderX

Data Exclusion

For the analyses regarding hypotheses 3.1 to 3.3, we excluded 10% (3/30) of the participants (3 women). One participant had technical problems with the headphones, which resulted in them performing the exercise twice without headphones. Another participant failed to disconnect the headphones, resulting in them performing the exercise twice with headphones. For the third participant, headphones could not capture any sound because of very silent breathing, whereas the smartphone microphone was able to capture some exhalations and missed most inhalations. We also excluded this third participant from the offline evaluation as the raters labeled most of the data from this participant as unclear. If we had included this participant, the analysis would have falsely shifted toward the hypotheses and arbitrarily favored the model's performance in the offline evaluation.

However, the data of these participants were included for all the other tests as the participants were still able to complete the 2 breathing training sessions, although the third participant received very erroneous breathing feedback. For the latter, the model predicted 99.46% of the headphone session and 93.53% of the smartphone session as nonbreathing because most of the captured sound was completely silent.

Ethics Approval

The Ethics Commission of the Swiss Federal Institute of Technology in Zurich (ID 2021-N-134) approved the study, and we pretested the study with 3 participants (1 woman).

Results

Checks for Reliability

We calculated reliability checks (Table 4) using the McDonald ω for the UES-SF and its subscales and for PE and perceived relaxation effectiveness (items 1 and 5 of the PE construct). The data from all these scales met the tests for normal distribution.

Table 4. Reliability tests for each survey construct.

Construct and subscale (number of items)	McDonald ω
User Engagement Score Short Form (12)	0.78
Focused attention (3)	0.53
Perceived usability (3)	0.58
Esthetic appeal (3)	0.79
Reward factor (3)	0.82
Perceived effectiveness (6)	0.58
Perceived relaxation effectiveness (2)	0.85

Hypothesis Tests

User Engagement

A Wilcoxon test indicated that the UES was higher than the neutral response (mean 3.77, SD 0.43) for the participants (W=459; P<.001). The difference was also observed for all the subscales: focused attention (mean 3.22, SD 0.66; W=245; adjusted P=.15), perceived usability$_{rev}$ (mean 3.90, SD 0.66; W=348; adjusted P<.001), esthetic appeal (mean 4.00, SD 0.547; W=406; adjusted P<.001), and reward factor (mean 3.97, SD 0.69; W=390; adjusted P<.001).

Effectiveness

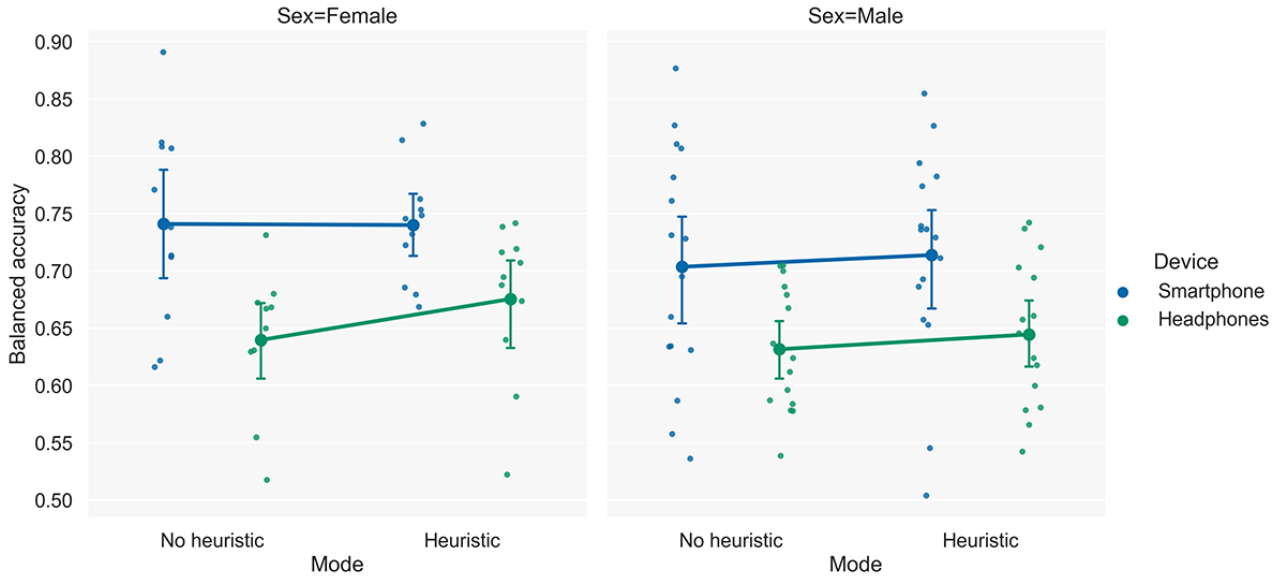
The reported PE was higher than the neutral response (mean 4.08, SD 0.49), as shown by a Wilcoxon test (W=465; adjusted P<.001). In addition, for perceived relaxation effectiveness (mean 3.82, SD 0.95), a Wilcoxon test indicated a positive effect (W=358; adjusted P<.001).

Breathing Detection Performance

A repeated-measure ANOVA indicated the presence of significant effects of headphone use ($F_{1,25}$=17.61; P<.001) and use of the heuristic ($F_{1,25}$=6.23; P=.02) on the detection performance of the model. The analysis did not indicate any interaction effects between the use of headphones and the use of the heuristic ($F_{1,25}$=3.39; P=.08). Furthermore, no interaction effects of sex were found with headphone use ($F_{1,25}$=0.11; P=.74), use of the heuristic ($F_{1,25}$=0.25; P=.62), or both ($F_{1,25}$=2.53; P=.12). In addition, no between-subject effects of sex were found ($F_{1,25}$=1.38; P=.25). Figure 9 illustrates the estimated marginal means.

XSL•FO
RenderX

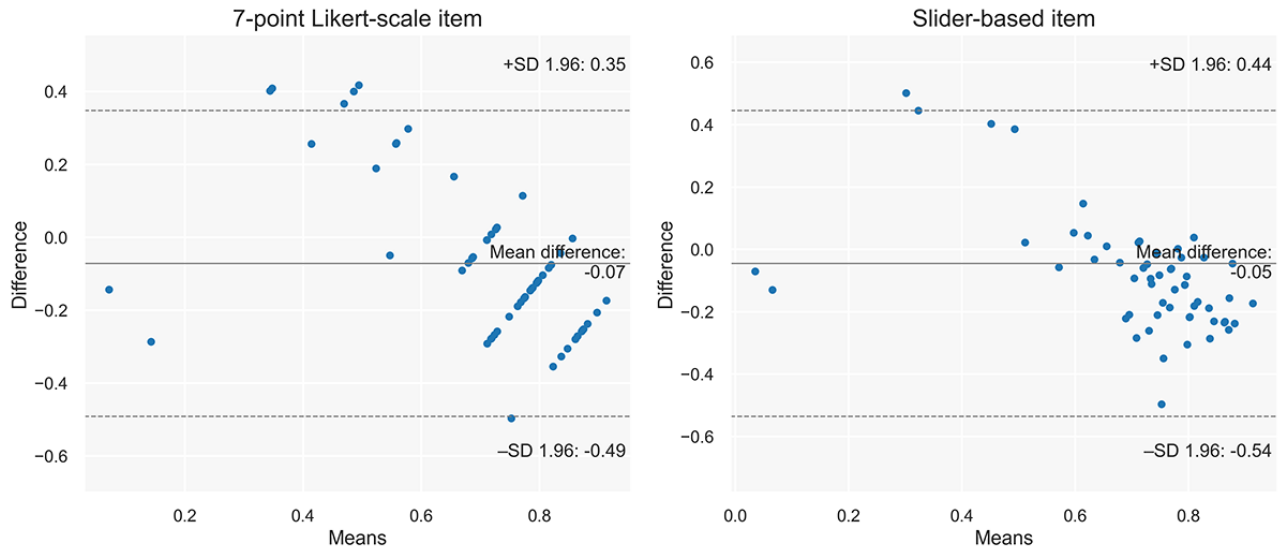
Figure 9. Marginal means plots illustrating the effects and interactions when different devices (smartphones and headphones) and model modes (no heuristic and heuristic) are used. Furthermore, the differences between female and male participants are also depicted.



Perceived Breathing Detection Performance

The perceived breathing detection accuracy reported via a 7-point Likert scale (mean 5.17, SD 1.75) and 0 to 100 slider (mean 71.17, SD 28.68) showed some correlation with the actual performance of the breathing detector (mean 0.69, SD 0.08). The Likert scale showed a stronger correlation (r=0.51; P<.001) with the actual detection performance than the perceived accuracy reported via the slider (r=0.48; P<.001). Nevertheless, both correlations were significant. Bland-Altman plots (Figure 10) for both items showed that, when actual breathing detection accuracy was low, participants overestimated the accuracy. At the same time, they underestimated the accuracy when the actual detection accuracy was high.

Figure 10. Bland-Altman plots for the 2 items measuring perceived breathing detection accuracy. Higher differences underestimate the actual detection accuracy, and lower values overestimate the detection accuracy. The limits of agreement are set to a 1.96 SD, which produces 95% CIs for the means of the differences.



Offline Breathing Detection Model Evaluation

The offline evaluation of the model (no heuristic) resulted in areas under the curve of 0.83, 0.87, and 0.91 for inhalation, exhalation, and nonbreathing sounds, respectively (Figure 11). The detailed results grouped by capturing device used on precision, recall/sensitivity, specificity, and balanced accuracy are shown in Table 5.

XSL•FO
RenderX

Figure 11. ROC curves (left) and confusion matrix (right) at the thresholds for breathing sounds that were used during the study. Both are calculated for the deployed model based on all data gathered throughout the breathing training sessions of the study participants. ROC: receiver operating characteristic.

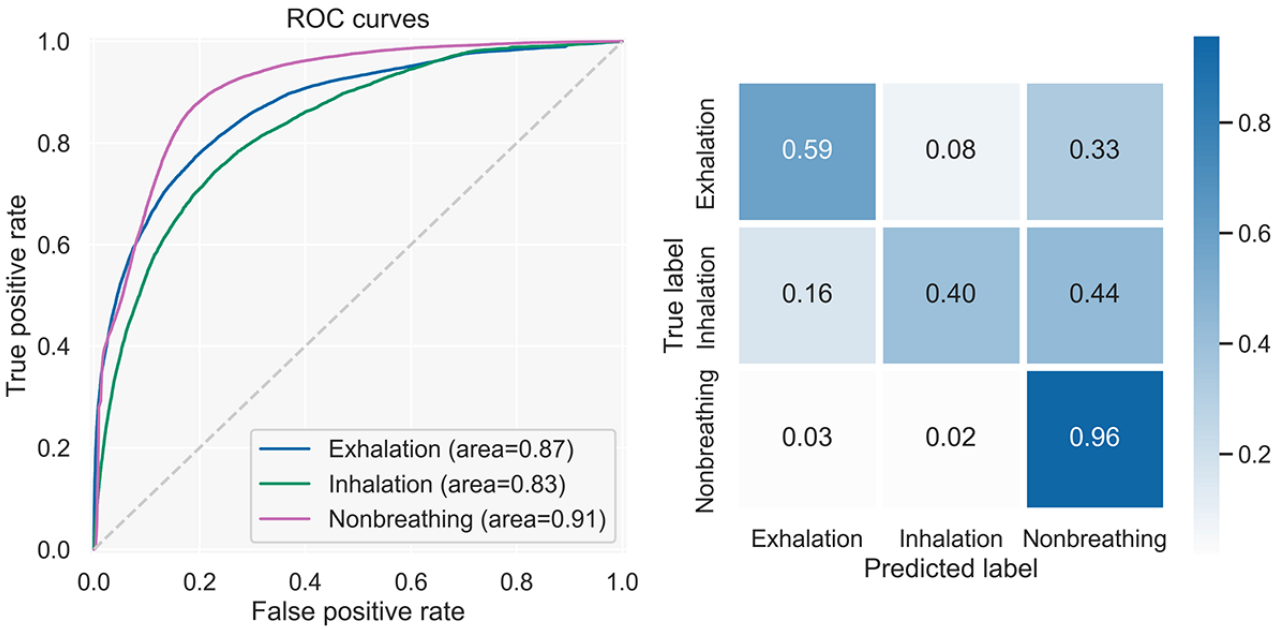


Table 5. Performance metrics of the model on the data captured during the study.

Class and device	Precision	Recall/sensitivity	Specificity	F_1 score	Balanced accuracy
Exhalation					
Combined	0.85	0.59	0.93	0.69	0.76
Smartphone	0.84	0.66	0.91	0.74	0.78
Headphones	0.86	0.51	0.94	0.64	0.73
Inhalation					
Combined	0.67	0.40	0.95	0.50	0.68
Smartphone	0.68	0.52	0.93	0.59	0.73
Headphones	0.65	0.26	0.97	0.37	0.61
Nonbreathing					
Combined	0.62	0.96	0.63	0.75	0.79
Smartphone	0.65	0.92	0.73	0.76	0.82
Headphones	0.59	0.99	0.53	0.74	0.76
All classes (average)					
Combined	0.71	0.65[a]	0.84	0.65	0.74[b]
Smartphone	0.72	0.70[a]	0.86	0.70	0.78[b]
Headphones	0.70	0.59[a]	0.81	0.59	0.70[b]

[a]Corresponds to multi-class balanced accuracy according to Kelleher et al [64].

[b]Corresponds to multi-class balanced accuracy according to Urbanowicz and Moore [58].

Discussion

Principal Findings

Overall, Breeze 2 was well received, and all 30 participants could handle all aspects of it. Furthermore, all participants (30/30, 100%) successfully performed two 3-minute sessions of breathing training.

The participants perceived Breeze 2 as engaging according to the UES that differed significantly from the neutral response. Thus, our data support hypothesis 1. The in-depth analysis of the focused attention, perceived usability, esthetic appeal, and reward factor subscales revealed that participants particularly

XSL•FO
RenderX

liked the esthetics and perceived reward factor of the experience, which were significantly higher than the neutral response. Interpretations of the focused attention and perceived usability scores were less conclusive than for the other 2 subscales owing to low reliability scores. Nevertheless, the average perceived usability was high and did differ significantly from the neutral response. However, focused attention was not significantly higher than the neutral response even though it had an average score with a positive tendency. This finding indicates that participants felt only moderately absorbed in the experience.

A feeling of absorption is important as this may lead to a flow state that helps people concentrate and perceive the task as rewarding and fun [65]. A possible solution for this could be to try out different modalities for breathing training such as virtual reality setups as such setups show promise for mindfulness exercises [24,66]. However, this would defeat the purpose of the objective regarding the intervention's scalability. Another approach could be to introduce more dominant short-term feedback loops [65] during training to foster immersion as the overall reward factor of the experience already appears to be high. Nevertheless, such feedback loops need to be implemented with care as too strong and exciting loops may counter the targeted effects of the training (eg, relaxation).

Our data also supported hypotheses 2.1 and 2.2, as PE and perceived relaxation effectiveness were significantly higher than the neutral response. However, although the reliability score for perceived relaxation effectiveness was high, it was relatively low for PE. Thus, the scores for the latter should be interpreted with caution. We argue that this low reliability could be because the PE scale includes the perceived relaxation effectiveness scale and several other items asking about not equally perceived aspects of the breathing training. For example, a few participants (5/30, 17%) did not feel relaxed by the training but still thought it was easy to follow the instructions and directed their attention to their breathing. This is supported by the fact that the perceived relaxation effectiveness subscale yielded high reliability while having a lower mean than the overall effectiveness scale. Nevertheless, the analyses support hypotheses 2 and 3, meaning that the participants overall regarded Breeze 2 as effective in guiding their breathing and, most importantly, inducing a feeling of relaxation. The results are in line with prior work [34,54].

The analysis regarding the impact of the heuristic on detection performance showed that the heuristic brought a significant improvement to the overall detection performance, thus supporting hypothesis 3.1. The use of headphone microphones instead of the built-in microphone of the smartphones had an even larger but negative effect on detection performance. This even larger negative effect supports hypothesis 3.2. We argue that there are 2 reasons for this. First, the use of the smartphone microphone allows the model to detect exhalations through the generated airflow. The initial model evaluation has shown that this works better than acoustic detection. Second, modern Bluetooth headphones are optimized for speech and, thus, use filters to reduce noise in audio signals (eg, the Apple AirPods second generation used [56]). Breathing sounds are very close to noise (eg, white noise) and, thus, trigger these reduction algorithms. Consequently, headphones may heavily suppress the breathing signal before the signal is passed to the model. How strongly these 2 reasons affect the observed negative effect remains unclear as data labeling did not differentiate between acoustically captured exhalations and exhalations captured through airflow. Regarding hypothesis 3.3, we did not observe any between-subject effects on detection performance based on the participants' sex.

In addition, our findings support hypothesis 4 as perceived breathing detection accuracy significantly correlated with the actual detection performance. Thus, perceived breathing detection accuracy appears to help capture how clearly the feedback is perceived and how well the algorithm performs. We observed that participants over- and underestimated the detector's performance when the actual performance was low and high, respectively. This over- and underestimation could indicate that the specific breathing feedback implementation in Breeze 2 gives users the sense of valid feedback even when the model performance is lacking. While conducting the study, we observed that participants felt more comfortable with the Likert-scale item than the slider-based item. Consequently, we plan to use the Likert-scale item in future studies to monitor perceived breathing detection accuracy in case changes need to be made to the feedback mechanism while Breeze 2 is deployed in the field.

Regarding the model without the heuristic, an apparent decline in detection performance was observed compared with the original test data set. The reasons for this are manifold. The breathing sound training and original test data sets were captured in the same setup [21]. This setup also differed considerably from the setup used in this study. In this study, the participants used Breeze 2 in a realistic setting for the first time by holding the device in their hands however was comfortable for them.

Consequently, the sound capturing was done in a much less regulated way. Furthermore, the training data used only a minimal number of devices shared between training and test data sets, whereas this study used smartphones and headphones unseen by the model. The headphones also pose a much more complex detection problem, as seen by the analysis regarding hypothesis 3.2. Our idea of a more complicated detection problem receives further support through the observation that all model performance metrics improve considerably for data only captured by smartphone microphones. This effect much more strongly influences the inhalation detection performance. We explain this through the observation that inhalations themselves are already very silent sounds and, thus, are already hard to detect. The attenuation applied by the headphones reinforces this problem even further.

Nevertheless, the model still performed reasonably well for exhalation sounds for both device types. The exhalation detection performance suffers, especially in sensitivity. However, the low sensitivity is less crucial as the model runs an inference up to 30 times per second, and the model yields high precision on exhalations. This reasoning is supported by the fact that participants overestimated the model's performance in most sessions (44/60, 73%), even when the model did perform poorly. Consequently, the model appears to be already usable

XSL•FO
RenderX

to enable interactivity in breathing training despite apparent weaknesses.

Limitations and Future Work

Optimizing Breeze is an iterative process and, consequently, it comes with several limitations. The slide-based tutorial is not very engaging and does not yet provide adequate guidance on breathing training details. We plan to improve the tutorial by providing video-based instructions. In addition, we are considering implementing an interactive tutorial to ensure that users can perform the breathing training correctly and give them feedback right away before they embark on an actual training session. Aside from the still too basic tutorial, Breeze 2 does not yet provide an adequate mechanism to coach the users on choosing the suitable training duration and breathing pattern apart from labeling the different durations according to their level of expertise with breathing training. We plan to develop interactive tests that measure users' capabilities and classify their level of experience (eg, based on the maximal time a person can inhale or exhale or on the user's resonance frequency that maximizes the physiological response). Such tests would then allow us to offer some coaching to the user on which training parameters would fit their level. Finally, we have planned various studies to incorporate Breeze 2 as an intervention component in multicomponent interventions (eg, interventions aiming to prevent NCDs, reduce distress in patients with cancer, or reduce acute stress in a student population).

This study also has some limitations regarding the detection model and its performance. A total of 2 raters did the labeling independently, and disagreements were carefully handled in a subsequent process. Consequently, confidence is high that the labels are correct. Nevertheless, 1.51% of the recordings were labeled as unclear in the resulting data set. Unclear parts were not used for further analysis, potentially leading to a slight overestimation of the performance of breathing detection. Furthermore, breathing can be very silent and, thus, may sometimes not be captured by microphones or may be actively suppressed by algorithms in the hardware used (eg, Bluetooth headphones). Therefore, the reported detection performance should be considered as the detection performance on breathing sounds that could be captured by the devices used. Furthermore, we trained the model on a minimal data set. Even though most participants perceived the model as performing well, there is room for improvement. With this study, we took the first step by collecting new data, which we will use to improve the model for future deployments.

Conclusions

This paper presented Breeze 2, a new iteration of the gameful breathing training app Breeze. It consists of a slow-paced breathing training guided by gameful visualizations and uses breathing-based interactions. Furthermore, it allows users to choose training parameters consisting of training duration and breathing pattern. These features should improve long-term adherence to breathing training, support individuals in doing breathing training correctly, and help continuously increase training intensity. To gain insight into whether Breeze 2 is engaging and effective and into the performance of the breathing detection used, we conducted a laboratory study with 30 participants. Results show that most participants (28/30, 93%) perceived Breeze 2 as engaging and effective.

Furthermore, breathing detection performed sufficiently well for most participants' sessions (50/60, 83%), as indicated by the perceived detection accuracy and actual detection accuracy. We attribute the exceptions to the combination of noise filtering done by the headphones and the very silent breathing of these participants, which was not audible in the recordings even to the raters conducting the labeling. We will use the collected breathing sounds to refine breathing detection, making it more stable and increasing its performance. Future work will use Breeze 2 as an intervention tool in various studies for the prevention and management of NCDs.

Acknowledgments

The authors would like to thank Helen Galliker for her support in the development of Breeze 2. Furthermore, the authors would like to thank Shari Klein for her valuable feedback and support during this study's analysis and writing phases. This study is cofunded by CSS Insurance, Switzerland. CSS Insurance had no role in the study design, app design, data management plans, or data analysis and interpretation of the results.

Authors' Contributions

YXL served as the principal developer of Breeze 2 and conceived the breathing detection approach. YXL and GWT designed and conducted the user study. TK provided advice regarding the development of Breeze 2 and the study design. YXL performed the statistical analyses with inputs from GWT. YXL wrote the report, while GWT and TK provided feedback on the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

YXL, GWT, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at Swiss Federal Institute of Technology Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. CSS was not involved in the study design, data collection, or analysis and interpretation of the results. EF and TK are also cofounders of Pathmate

XSL•FO
RenderX

Technologies, a university spin-off company that creates and delivers digital clinical pathways. Pathmate Technologies is not involved in the study app described in this paper.

Multimedia Appendix 1
Screen recording of a breathing training session with Breeze 2.
[MOV File , 188753 KB - games_v10i3e39186_app1.mov]

Multimedia Appendix 2
Questionnaire used in the user study.
[PDF File (Adobe PDF File), 12093 KB - games_v10i3e39186_app2.pdf]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Newman D, Tong M, Levine E, Kishore S. Prevalence of multiple chronic conditions by U.S. state and territory, 2017. PLoS One 2020 May 5;15(5):e0232346 [FREE Full text] [doi: 10.1371/journal.pone.0232346] [Medline: 32369509]
3. Murphy A, Palafox B, Walli-Attaei M, Powell-Jackson T, Rangarajan S, Alhabib KF, et al. The household economic burden of non-communicable diseases in 18 countries. BMJ Glob Health 2020 Feb 11;5(2):e002040 [FREE Full text] [doi: 10.1136/bmjgh-2019-002040] [Medline: 32133191]
4. Gevirtz R. The promise of heart rate variability biofeedback: evidence-based application. Biofeedback 2013;41(3):110-120. [doi: 10.5298/1081-5937-41.3.01]
5. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? Front Psychol 2014 Jul 21;5:756 [FREE Full text] [doi: 10.3389/fpsyg.2014.00756] [Medline: 25101026]
6. Russell ME, Scott AB, Boggero IA, Carlson CR. Inclusion of a rest period in diaphragmatic breathing increases high frequency heart rate variability: implications for behavioral therapy. Psychophysiology 2017 Mar;54(3):358-365 [FREE Full text] [doi: 10.1111/psyp.12791] [Medline: 27925652]
7. Schumer MC, Lindsay EK, Creswell JD. Brief mindfulness training for negative affectivity: a systematic review and meta-analysis. J Consult Clin Psychol 2018 Jul;86(7):569-583 [FREE Full text] [doi: 10.1037/ccp0000324] [Medline: 29939051]
8. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol Med 2017 Nov;47(15):2578-2586. [doi: 10.1017/S0033291717001003] [Medline: 28478782]
9. Siepmann M, Aykac V, Unterdörfer J, Petrowski K, Mueck-Weymann M. A pilot study on the effects of heart rate variability biofeedback in patients with depression and in healthy subjects. Appl Psychophysiol Biofeedback 2008 Dec;33(4):195-201. [doi: 10.1007/s10484-008-9064-z] [Medline: 18807175]
10. Wang MH, Yeh ML. Respiratory training interventions improve health status of heart failure patients: a systematic review and network meta-analysis of randomized controlled trials. World J Clin Cases 2019 Sep 26;7(18):2760-2775 [FREE Full text] [doi: 10.12998/wjcc.v7.i18.2760] [Medline: 31616691]
11. Mendes LP, Moraes KS, Hoffman M, Vieira DS, Ribeiro-Samora GA, Lage SM, et al. Effects of diaphragmatic breathing with and without pursed-lips breathing in subjects with COPD. Respir Care 2019 Feb;64(2):136-144 [FREE Full text] [doi: 10.4187/respcare.06319] [Medline: 30154127]
12. Russo MA, Santarelli DM, O'Rourke D. The physiological effects of slow breathing in the healthy human. Breathe (Sheff) 2017 Dec;13(4):298-309 [FREE Full text] [doi: 10.1183/20734735.009817] [Medline: 29209423]
13. Magnon V, Dutheil F, Vallet GT. Benefits from one session of deep and slow breathing on vagal tone and anxiety in young and older adults. Sci Rep 2021 Sep 29;11(1):19267 [FREE Full text] [doi: 10.1038/s41598-021-98736-9] [Medline: 34588511]
14. Lin IM, Tai LY, Fan SY. Breathing at a rate of 5.5 breaths per minute with equal inhalation-to-exhalation ratio increases heart rate variability. Int J Psychophysiol 2014 Mar;91(3):206-211. [doi: 10.1016/j.ijpsycho.2013.12.006] [Medline: 24380741]
15. Porges SW. The polyvagal perspective. Biol Psychol 2007 Feb;74(2):116-143 [FREE Full text] [doi: 10.1016/j.biopsycho.2006.06.009] [Medline: 17049418]
16. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
17. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]

XSL•FO
RenderX

18. Dicheva D, Dichev C, Agre G, Angelova G. Gamification in education: a systematic mapping study. Educ Technol Soc 2015;18(3):75-88.
19. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov 2;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
20. Pham Q, Khatib Y, Stansfeld S, Fox S, Green T. Feasibility and efficacy of an mHealth game for managing anxiety: "Flowy" randomized controlled pilot trial and design evaluation. Games Health J 2016 Feb;5(1):50-67. [doi: 10.1089/g4h.2015.0033] [Medline: 26536488]
21. Shih CH, Tomita N, Lukic YX, Reguera Á, Fleisch E, Kowatsch T. Breeze: smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 11;3(4):1-30. [doi: 10.1145/3369835]
22. Agrawal V, Naik V, Duggirala M, Athavale S. Calm a mobile based deep breathing game with biofeedback. In: Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play. 2020 Nov Presented at: CHI PLAY '20; November 2-4, 2020; Virtual p. 153-157. [doi: 10.1145/3383668.3419876]
23. Tabor A, Pradantyo R, Sadprasid B, Birk MV, Scheme E, Bateman S. Bubble breather-a breathing exercise game to support pneumonia rehabilitation and recovery. In: Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play. 2020 Nov Presented at: CHI PLAY '20; November 2-4, 2020; Virtual p. 86-90. [doi: 10.1145/3383668.3419921]
24. Patibanda R, Mueller FF, Leskovsek M, Duckworth J. Life tree: understanding the design of breathing exercise games. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY '17; October 15-18, 2017; Amsterdam, The Netherlands p. 19-31. [doi: 10.1145/3116595.3116621]
25. Weerdmeester J, van Rooij MM, Maciejewski DF, Engels RC, Granic I. A randomized controlled trial assessing the efficacy of a virtual reality biofeedback video game: anxiety outcomes and appraisal processes. Technol Mind Behav 2021 Aug 05;2(2):tmb0000028. [doi: 10.1037/tmb0000028]
26. Delk KK, Gevirtz R, Hicks DA, Carden F, Rucker R. The effects of biofeedback assisted breathing retraining on lung functions in patients with cystic fibrosis. Chest 1994 Jan;105(1):23-28. [doi: 10.1378/chest.105.1.23] [Medline: 8275737]
27. Kaushik R, Kaushik RM, Mahajan SK, Rajesh V. Biofeedback assisted diaphragmatic breathing and systematic relaxation versus propranolol in long term prophylaxis of migraine. Complement Ther Med 2005 Sep;13(3):165-174. [doi: 10.1016/j.ctim.2005.04.004] [Medline: 16150370]
28. Thiebes S, Lins S, Basten D. Gamifying information systems-a synthesis of gamification mechanics and dynamics. In: Proceedings of the 22nd European Conference on Information Systems. 2014 Presented at: ECIS '14; June 9-11, 2014; Tel Aviv, Israel p. 1-17.
29. Liu D, Santhanam R, Webster J. Toward meaningful engagement: a framework for design and research of gamified information systems. MIS Q 2017 Apr 4;41(4):1011-1034. [doi: 10.25300/misq/2017/41.4.01]
30. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of brief biofeedback via a smartphone app on stress recovery: randomized experimental study. JMIR Serious Games 2019 Nov 26;7(4):e15974 [FREE Full text] [doi: 10.2196/15974] [Medline: 31769761]
31. Faust-Christmann CA, Taetz B, Zolynski G, Zimmermann T, Bleser G. A biofeedback app to instruct abdominal breathing (breathing-mentor): pilot experiment. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13703 [FREE Full text] [doi: 10.2196/13703] [Medline: 31573919]
32. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
33. Lukic YX, Shih CH, Hernandez Reguera A, Cotti A, Fleisch E, Kowatsch T. Physiological responses and user feedback on a gameful breathing training app: within-subject experiment. JMIR Serious Games 2021 Feb 08;9(1):e22802 [FREE Full text] [doi: 10.2196/22802] [Medline: 33555264]
34. Lukic YX, Klein SS, Brügger V, Keller OC, Fleisch E, Kowatsch T. The impact of a gameful breathing training visualization on intrinsic experiential value, perceived effectiveness, and engagement intentions: between-subject online experiment. JMIR Serious Games 2021 Sep 14;9(3):e22803 [FREE Full text] [doi: 10.2196/22803] [Medline: 34519662]
35. Islam B, Rahman MM, Ahmed T, Ahmed MY, Hasan MM, Nathan V, et al. BreathTrack: detecting regular breathing phases from unannotated acoustic data captured by a smartphone. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Sep 09;5(3):1-22. [doi: 10.1145/3478123]
36. Fischer T, Schneider J, Stork W. Classification of breath and snore sounds using audio data recorded with smartphones in the home environment. In: Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. 2016 Presented at: ICASSP '16; March 20-25, 2016; Shanghai, China p. 226-230. [doi: 10.1109/icassp.2016.7471670]
37. Romero HE, Ma N, Brown GJ, Beeston AV, Hasan M. Deep learning features for robust detection of acoustic events in sleep-disordered breathing. In: Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. 2019 Presented at: ICASSP '19; May 12-17, 2019; Brighton, UK p. 810-814. [doi: 10.1109/icassp.2019.8683099]

XSL•FO
RenderX

38. Ren Y, Wang C, Yang J, Chen Y. Fine-grained sleep monitoring: hearing your breathing with smartphones. In: Proceedings of the 2015 IEEE Conference on Computer Communications. 2015 Presented at: INFOCOM '15; April 26-May 1, 2015; Hong Kong, China p. 1194-1202. [doi: 10.1109/infocom.2015.7218494]
39. Kumar A, Mitra V, Oliver C, Ullal A, Biddulph M, Mance I. Estimating respiratory rate from breath audio obtained through wearable microphones. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:7310-7315. [doi: 10.1109/EMBC46164.2021.9629661] [Medline: 34892786]
40. Plakal M, Ellis D. Tensor Flow - models: YAMNet. GitHub. 2021 Sep 9. URL: https://github.com/tensorflow/models/tree/master/research/audioset/yamnet [accessed 2021-08-20]
41. Freesound. URL: https://freesound.org/ [accessed 2022-08-10]
42. Piczak KJ. ESC: dataset for environmental sound classification. In: Proceedings of the 23rd ACM international conference on Multimedia. 2015 Presented at: MM '15; October 26-30, 2015; Brisbane, Australia p. 1015-1018. [doi: 10.1145/2733373.2806390]
43. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv 2022 Preprint posted online on April 17, 2017. [FREE Full text] [doi: 10.48550/arXiv.1704.04861]
44. Gemmeke JF, Ellis DP, Freedman D, Jansen A, Lawrence W, Moore RC, et al. Audio set: an ontology and human-labeled dataset for audio events. In: Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. 2017 Presented at: ICASSP '17; March 5-9, 2017; New Orleans, LA, USA p. 776-780. [doi: 10.1109/icassp.2017.7952261]
45. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021 Jan;109(1):43-76. [doi: 10.1109/jproc.2020.3004555]
46. Ng A, Lepinski J, Wigdor D, Sanders S, Dietz P. Designing for low-latency direct-touch input. In: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology. 2012 Presented at: UIST '12; October 7-10, 2012; Cambridge, MA, USA p. 453-464. [doi: 10.1145/2380116.2380174]
47. Deber J, Jota R, Forlines C, Wigdor D. How Much Faster is Fast Enough?: User Perception of Latency & Latency Improvements in Direct and Indirect Touch. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Apr Presented at: CHI '15; April 18-23, 2015; Seoul, South Korea p. 1827-1836. [doi: 10.1145/2702123.2702300]
48. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(11):2579-2605.
49. Ramachandran P, Zoph B, Le QV. Searching for activation functions. arXiv 2022 Preprint posted online on October 27, 2017. [FREE Full text] [doi: 10.48550/arXiv.1710.05941]
50. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
51. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. 2016 Presented at: OSDI '16; November 2-4, 2016; Savannah, GA, USA.
52. LoMauro A, Aliverti A. Sex differences in respiratory function. Breathe (Sheff) 2018 Jun;14(2):131-140 [FREE Full text] [doi: 10.1183/20734735.000318] [Medline: 29875832]
53. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Human Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
54. Chittaro L, Sioni R. Evaluating mobile apps for breathing training: the effectiveness of visualization. Comput Human Behav 2014 Nov;40:56-63. [doi: 10.1016/j.chb.2014.07.049]
55. Efendić E, Van de Calseyde PP, Evans AM. Slow response times undermine trust in algorithmic (but not human) predictions. Organ Behav Hum Decis Process 2020 Mar;157:103-114. [doi: 10.1016/j.obhdp.2020.01.008]
56. AirPods (2nd generation) with Charging Case - Technical Specifications. Apple Inc. 2021. URL: https://support.apple.com/kb/SP792 [accessed 2022-03-10]
57. McDonald RP. Test Theory: A Unified Treatment. London, UK: Psychology Press; 2013.
58. Urbanowicz RJ, Moore JH. ExSTraCS 2.0: description and evaluation of a scalable learning classifier system. Evol Intell 2015 Sep;8(2):89-116 [FREE Full text] [doi: 10.1007/s12065-015-0128-8] [Medline: 26417393]
59. Velez DR, White BC, Motsinger AA, Bush WS, Ritchie MD, Williams SM, et al. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol 2007 May;31(4):306-315. [doi: 10.1002/gepi.20211] [Medline: 17323372]
60. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965 Dec 01;52(3/4):591-611. [doi: 10.2307/2333709]
61. Olkin I. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford, CA, USA: Stanford University Press; 1960.
62. Brown MB, Forsythe AB. Robust tests for the equality of variances. J Am Stat Assoc 1974 Jun;69(346):364-367. [doi: 10.1080/01621459.1974.10482955]

XSL•FO
RenderX

63. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
64. Kelleher JD, Mac Namee B, D'arcy A. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, MA, USA: MIT press; 2015.
65. Cowley B, Charles D, Black M, Hickey R. Toward an understanding of flow in video games. Comput Entertain 2008 Jul;6(2):1-27. [doi: 10.1145/1371216.1371223]
66. Roo JS, Gervais R, Frey J, Hachet M. Inner garden: connecting inner states to a mixed reality sandbox for mindfulness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO, USA p. 1459-1470. [doi: 10.1145/3025453.3025743]

Abbreviations

BPM: breaths per minute
HRV: heart rate variability
NCD: noncommunicable disease
PE: perceived effectiveness
ROC: receiver operating characteristic
UES: user engagement score
UES-SF: User Engagement Scale Short Form
YAMNet: Yet Another Mobile Network

Edited by N Zary; submitted 06.05.22; peer-reviewed by H Li, A Teles, N Chalghaf; comments to author 30.05.22; revised version received 28.06.22; accepted 21.07.22; published 16.08.22.

Please cite as:
Lukic YX, Teepe GW, Fleisch E, Kowatsch T
Breathing as an Input Modality in a Gameful Breathing Training App (Breeze 2): Development and Evaluation Study
JMIR Serious Games 2022;10(3):e39186
URL: https://games.jmir.org/2022/3/e39186
doi:10.2196/39186
PMID:35972793

©Yanick Xavier Lukic, Gisbert Wilhelm Teepe, Elgar Fleisch, Tobias Kowatsch. Originally published in JMIR Serious Games (https://games.jmir.org), 16.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Therapeutic Video Game With the MDA Framework to Decrease Anxiety in Preschool-Aged Children With Acute Lymphoblastic Leukemia: Mixed Methods Approach

Dai-Jie Yang[1,2], RN, MSN; Meng-Yao Lu[3], MD; Chi-Wen Chen[1], RN, PhD; Pei-Ching Liu[1], RN, PhD; I-Ching Hou[1,4], RN, PhD

[1]College of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan

[2]Department of Nursing, National Taiwan University Hospital, Taipei City, Taiwan

[3]Department of Pediatrics, National Taiwan University Hospital, Taipei City, Taiwan

[4]Efficient Smart Care Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
I-Ching Hou, RN, PhD
College of Nursing
National Yang Ming Chiao Tung University
Number 155, Section 2, Linong Street
Beitou District
Taipei City, 11221
Taiwan
Phone: 886 28267000 ext 67315
Email: evidta@gmail.com

Related Article:

This is a corrected version. See correction statement: https://games.jmir.org/2022/4/e43211

Abstract

Background: Preschool-aged children with acute lymphoblastic leukemia (ALL) receive long-term treatment according to the Taiwan Pediatric Oncology Group (TPOG)–ALL 2013 protocol. Severe anxiety and noncompliance ahead of frequent invasive therapies leads to an increase in health care costs. Previous studies have shown that therapeutic video games (TVGs) can decrease the anxiety experienced by children who are ill. To our knowledge, no existing TVG has been designed specifically for preschool-aged children with ALL in Taiwan.

Objective: The purpose of this study was to develop a TVG using the popular Mechanics, Dynamics, and Aesthetics (MDA) framework for game design and to investigate the effect of this TVG on the reduction of therapy-related anxiety among preschool-aged children with ALL.

Methods: This study used a mixed methods approach over three phases: (1) develop a TVG using the MDA framework, (2) test the reliability of the TVG among three certified children's art therapists, and (3) evaluate the reduction of therapy-related anxiety among participants after using the TVG for 6 weeks, using a two-group, stratified randomized controlled trial at a medical center in northern Taiwan. Eligible preschool-aged children with ALL were randomly assigned 1:1 into an experimental group or a control group. The two groups of subjects received the same usual care, and only the experimental group had access to and used the TVG. The children's anxiety responses were reported by their family caregivers using the face rating scale (FRS). Descriptive analyses, the Fisher exact test, the Pearson chi-square test, and the Mann-Whitney U test were used to statistically analyze the variables.

Results: Six mechanics rules supported the dynamics of the TVG using four main features—character, nursery, tasks, and market—in order to complete all of the therapy-related anxiety reduction scenarios and to achieve eight aesthetics goals. The results of reliability test showed that participants found the TVG to be useful and trustworthy for preschool-aged children with ALL (Cronbach α=.98). A total of 15 participants were enrolled and randomly allocated to the experimental group (n=7) or the control group (n=8). The average number of TVG log-ins was 37.9 (SD 15.30, range 14-62) in the experimental group. The demographic data showed homogeneity across the two groups regarding age (3 to 5 years), sex (male), risk classification (standard

XSL•FO
RenderX

risk), and treatment status (continuation therapy). The mean FRS score was 6.16 (SD 3.31) for the experimental group as compared to 7.45 (SD 2.71) for the control group (P=.04), which represented a significant difference between the groups at the 6-week follow-up.

Conclusions: This research provides evidence that using a TVG can decrease anxiety in preschool-aged children with ALL in Taiwan. The TVG could be used to support clinical professionals before they perform invasive therapies. However, it is recommended to increase the statistical power for inference.

Trial Registration: ClinicalTrials.gov NCT04199637; https://www.clinicaltrials.gov/ct2/show/NCT04199637

(JMIR Serious Games 2022;10(3):e37079) doi:10.2196/37079

KEYWORDS

acute lymphoblastic leukemia; therapeutic video games; childhood cancer; preschoolers; anxiety

Introduction

Background

Acute lymphoblastic leukemia (ALL) is a type of cancer that constitutes a family of genetically heterogeneous lymphoid neoplasms derived from B- and T-lymphoid progenitors [1]; this type of cancer in humans shows an abnormal increase of leukocytes in the blood. ALL occurs mostly in children, particularly in those between 2 and 5 years of age, and its prevalence rate is about 30% to 40% in children worldwide [2]. In Taiwan, approximately 100 children aged 1 to 9 years are diagnosed with ALL every year, and more than half of them are preschoolers (ie, 3-5 years old) [3].

To obtain a definite diagnosis of ALL, bone marrow aspiration (BMA) must be performed. This is an invasive procedure performed by a physician to collect a small sample of bone marrow from a child's hip bone, breastbone, or thigh bone under local anesthesia. Meanwhile, other body samples are also collected, such as peripheral blood cells, cerebrospinal fluid, and testicular cells. These results support the differential diagnosis of ALL, including the following: precursor B-cell ALL, early T-cell precursor (ETP) ALL, and precursor T-cell ALL with non-ETP [4].

The Taiwan Pediatric Oncology Group (TPOG), which consists of more than 60 pediatric hematologists and oncologists, classified children with ALL into one of three categories: standard risk, high risk, or very high risk. Classifications were based on the children's presenting age, leukocyte count, presence of leukemia cells in the cerebrospinal fluid or testicular leukemia, immunophenotype, cytogenetics and molecular genetics, DNA index, and early response to therapy [2,4,5]. However, the risk categories will be reconfirmed according to the response to the protocol based on the minimal residual disease, which refers to a small number of cancer cells that remain in the body after treatment [6].

The TPOG-ALL 2013 Protocol

The TPOG developed the first protocol in 1988 for childhood ALL; the protocol has since had several revisions [7]. At the beginning of our study, the TPOG-ALL 2013 protocol was used to treat children in Taiwan over four regimen stages: induction, consolidation, reinduction, and continuation. In this protocol, each stage consisted of different doses of antileukemic drugs (eg, prednisolone, vincristine, epirubicin, L-asparaginase, cyclophosphamide, cytarabine, and 6-mercaptopurine). The drugs were administrated by health professionals (eg, physicians and nurses) via intravenous (IV) injection, intramuscular (IM) injection, intrathecal (IT) injection, and oral intake. Multimedia Appendix 1 shows the summary of the protocol.

To facilitate the IV injection procedure, an artificial blood vessel, using a port-a-cath catheter system (PORT), would be implanted into the children's left or right superior vena cava via surgery before the protocol. The needle with catheter—usually the Huber-point needles are used—would be inserted before IV injection and reinserted according to the clinical routine (eg, every 3 to 7 days) to prevent potential infection. The most common clinical signs and symptoms (ie, side effects) during the treatment were nausea, vomiting, mucositis, hair loss, bone marrow suppression, and pain [8]. Therefore, supportive medication and blood transfusions (BTs) will be provided as needed intravenously. In addition, to assess antileukemic response, a BMA will also be executed following the protocol, and the minimal residual disease will be monitored to guide the therapeutic choices [6].

Therapy-Related Anxiety

A child's ALL diagnosis can undoubtedly have a major impact on their family. Current intensive chemotherapy regimens have resulted in overall cure rates of 85% to 90% in children [1,7]. However, the discomfort caused by the side effects of the treatments [8] and the series of invasive therapies that must be endured during the disease journey all increase the anxiety of the children suffering from disease [9]. They also might resist professional treatment, leading to degeneration and overdependence behaviors [10].

In order to reduce therapy-related anxiety of children who are ill, many medical professionals have used cognitive and attentional distraction in order to control the anxiety of pediatric patients before, during, and after treatments [11,12]. However, medical staff have spent lots of time accompanying children who were ill, which was an invisible time cost, and the demand for clinical and medical labor has increased. Electronic technology products, such as computers, smartphones, and tablets, have become popular therapeutic adjuncts to traditional approaches (eg, drawing, role play, and toy operations). The games in portable electronic products have gradually become another aspect of children's game activities. Many studies have applied video games or virtual reality to pediatric care [13-15]. However, this type of research is mostly limited to the use of

XSL•FO
RenderX

general electronic games to divert attention, rather than specially designed therapeutic electronic games.

It has been clinically observed that the frequency of use of portable electronic products by preschool-aged children is extremely high. Almost every family caregiver will provide portable electronic products to children who are ill so they may watch videos. This shows that the development of related game products may be helpful for alleviating the anxiety experienced by children who are ill and may help them understand the purpose of treatment. Therefore, the purpose of this study was to develop a video game that can be played by preschool-aged children with ALL based on their cognitive development and game patterns, in order to understand how therapeutic video games (TVGs) can improve the anxiety response experienced by these children.

The Mechanics, Dynamics, and Aesthetics Framework

The Mechanics, Dynamics, and Aesthetics (MDA) framework for game design and research was published by Hunicke et al in 2004 [16]. Mechanics describes the particular components of the game, at the level of data representation and algorithms (eg, rules). Dynamics describes the run-time behavior of the mechanics acting on the player inputs and each other's outputs over time. Aesthetics describes the desirable emotional responses evoked in the player when the player interacts with the game system. These emotional responses include, but are not limited to, the taxonomy of sensation (eg, game as sense-pleasure), fantasy (eg, game as make-believe), narrative (eg, game as drama), challenge (eg, game as obstacle), fellowship (eg, game as social framework), discovery (eg, game as uncharted territory), expression (eg, game as self-discovery), and submission (eg, game as pastime). Each component of the MDA framework is causally linked by the game designers and provides the players with pleasure and fun perceptions [16]. When designing video games for preschool-aged children who are ill in order to decrease their therapy-related anxiety, the MDA framework could act as the guideline for the game designer.

Methods

Study Design

In this study, which began in 2018, our team conducted a mixed methods approach with three phases. In the first phase, the MDA framework [16] was adopted to design a TVG. In the second phase, a reliability test was applied before the TVG was used. In the third phase, a randomized controlled trial (RCT) was conducted to evaluate therapy-related anxiety experienced by preschool-aged children with ALL during their treatment according to the TPOG-ALL 2013 protocol. This study was registered at ClinicalTrials.gov (NCT04199637).

Phase 1: Design of a TVG With the MDA Framework

To decrease therapy-related anxiety experienced by preschool-aged children with ALL, the research team cooperated with a Taiwan-based video game technology company to develop a TVG. Group meetings were held to reach a consensus regarding the MDA framework [16]. The game designer (DJY), who was also a registered nurse in the pediatric oncology ward, expressed that based on the knowledge of preschool-aged children's cognitive-developmental stage (eg, preoperational), children 3 to 5 years of age show egocentrism, show anthropomorphism, think without logic, and lack conservation [17]. Thus, to facilitate aesthetics when playing the game, the TVG design should avoid mechanics and dynamics with complex logical (eg, calculation) or conservation (eg, comparison) aspects, but should provide an anthropomorphic role, easy and repeatable activities, and low-challenge tasks.

TVG development consensus meetings were held regularly (eg, once per week). The team first focused on the aesthetics expectations and dynamic features and then defined the mechanics data representations and algorithms. When the MDA framework was conceptually saturated and logically feasible, the game designer (DJY) provided a user interface draft for our TVG, including procedures for invasive therapies and required medical supplies (Multimedia Appendix 2), to the visual designer and programmer to help them design the game for preschool-aged children with ALL. Once the prototype and vital version of the TVG was produced, the research team tested and retested all the features, after which it was assessed during phase 2.

Phase 2: TVG Reliability Test

To assess the reliability of the TVG, three certified children's art therapists were invited to preview the TVG. After previewing the game, they completed a 12-item instrument containing two domains—usefulness (9 items) and trust (2 items)—with structured questions answered on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A single open question allowed input of narrative feedback about the game. We performed the following to evaluate the usefulness and participants' trust of the TVG: descriptive analysis of the 11 structured items, content analysis of the single narrative feedback item, and internal consistency (ie, Cronbach α) of the three therapists, in order to analyze reliability. The results were also used to optimize the TVG before phase 3.

Phase 3: RCT to Measure the Effects of the TVG

Study Design and Setting

The TVG trial was a patient-blinded, parallel-group, stratified RCT with a 6-week follow-up evaluation. The study site was the medical center (eg, National Taiwan University Hospital) in northern Taiwan. Every patient at the study site received the same verbal comfort from nurses before and during the process of the invasive therapy. Some attention-shifting skills (eg, watching cartoons, listening to nursery rhymes and kids' songs, and being given comforting toys) were also adopted, but these were nonstandardized and enabled by their caregivers and professionals to facilitate the therapy procedure. Therefore, the influence of the professionals' care model on anxiety improvement was assumed to be similar for each patient in this study.

Ethical Considerations

This study followed the ethical principles of the Declaration of Helsinki [18]. This study was approved by the Institutional Review Board (IRB) of the National Taiwan University Hospital

XSL•FO
RenderX

(IRB No. 201705014RINC) in Taiwan. Participation in the study was voluntary. Eligible participants were provided with one of two informed consent forms, according to their assigned groups. The consent forms included the same information about participant data that were collected regarding therapy-related anxiety responses. Only the consent forms for the experimental group provided briefing information about the TVG. Every participant was provided US $0.70 for each follow-up evaluation during the study.

Participant Recruitment, Sample Size, and Randomization

Preschool-aged children who met the following inclusion criteria were invited to participate in the study: (1) diagnosis of ALL and (2) aged 3 to 5 years. Exclusion criteria included the following: (1) has not undergone treatment according to the TPOG-ALL 2013 protocol, (2) did not have the PORT system inserted, (3) undergoing peripheral blood stem cell transplant treatment, (4) experiencing recurrent ALL, and (5) diagnosis of mental retardation, because of which TVG would not support their needs in this trial.

Participants were recruited by the researcher (DJY) and one trained research assistant in the pediatric hematology ward or outpatient clinic. Eligible patients were screened by the researcher (DJY) before recruitment in the study setting, according to the 2015-2016 Taiwan ALL incidence rate within each age and sex strata [3]. Next, within each age and sex strata, participants were randomly assigned 1:1 into one of the two study groups: control or experimental. For each strata, the randomization scheme was generated using an online randomization program [19].

Data Collection Procedures

Data collection by the study team took place from March 2018 to April 2020. Patients were blinded as to whether they were in the experimental or control group. Demographics data (eg, age, sex, diagnosis, risk classifications, and treatment stage) of the eligible participants were collected through the electronic health record at the study site. The TVG was only installed on the smartphones of the experimental group participants' primary caregivers; they were then taught how to use it. Each participant in the experimental group could play the TVG at any time as needed (eg, no prompts or reminders from the study team were needed). To ensure that the patients in the control group were not receiving the intervention, the TVG web address and log-in account details were only provided by the research team who were, themselves, not able to access the game. Meanwhile, the experimental group caregivers were told, and agreed to, not to share the TVG log-in account details with any other children with ALL during the trial.

The 6-week follow-up evaluation data collection was completed via a web-based instrument following the invasive therapy. To prevent missing records, communication software (eg, Line) was used to remind the participant caregivers every week. All of the intervention group participants received the same TVG (eg, frozen version), and there were no technical problems associated with the TVG during the trial.

Therapy-Related Anxiety Response Report

In this study, therapy-related anxiety responses by the children were reported by their family caregivers through a web form that was installed on their mobile phones at enrollment. This form consisted of four items: invasive therapy date, invasive therapies received, face rating scale (FRS) score, and other responses. The second item—invasive therapies conducted—included six activities: buttocks injection, chemical medication administration, IT injection, BMA, PORT puncture, and blood transfusion. The third item–the FRS–included 10 faces with corresponding numbers: a smiling face with mouth open (ie, number 1) indicates not crying at all and a red face with two tears (ie, number 10) indicates severe crying. The fourth item—other responses—allowed family caregivers to note any observations about the participants in a free-text field. When the participant completed any one of six invasive therapies, the family caregiver would help them point out on the FRS the face that expressed their real feelings during that therapy. The average time that the caregivers spent completing the report was less than 5 minutes without any assistance.

Data Analysis

All reported data were coded and analyzed using SPSS (version 24; IBM Corp). Frequency and percentage as well as mean and SD were used as descriptive statistics for the demographic data and for FRS results. We compared the age, sex, risk classifications, treatment status, invasive therapies received, and FRS scores of participants in the control and experimental groups; the Fisher exact test and Pearson chi-square tests were used for categorical variables, and the Mann-Whitney *U* test was used for continuous variables. Significance was defined as $P<.05$. For the other responses, each report was categorized as positive or negative, which was defined by the research team with consensus. Positive responses were defined as the participants not crying or being just a little nervous at the beginning of the therapy. Negative responses were defined as the participants crying, reacting physically (eg, vomiting), or showing any resistance behaviors (eg, trying to run away).

Results

The Content of the TVG

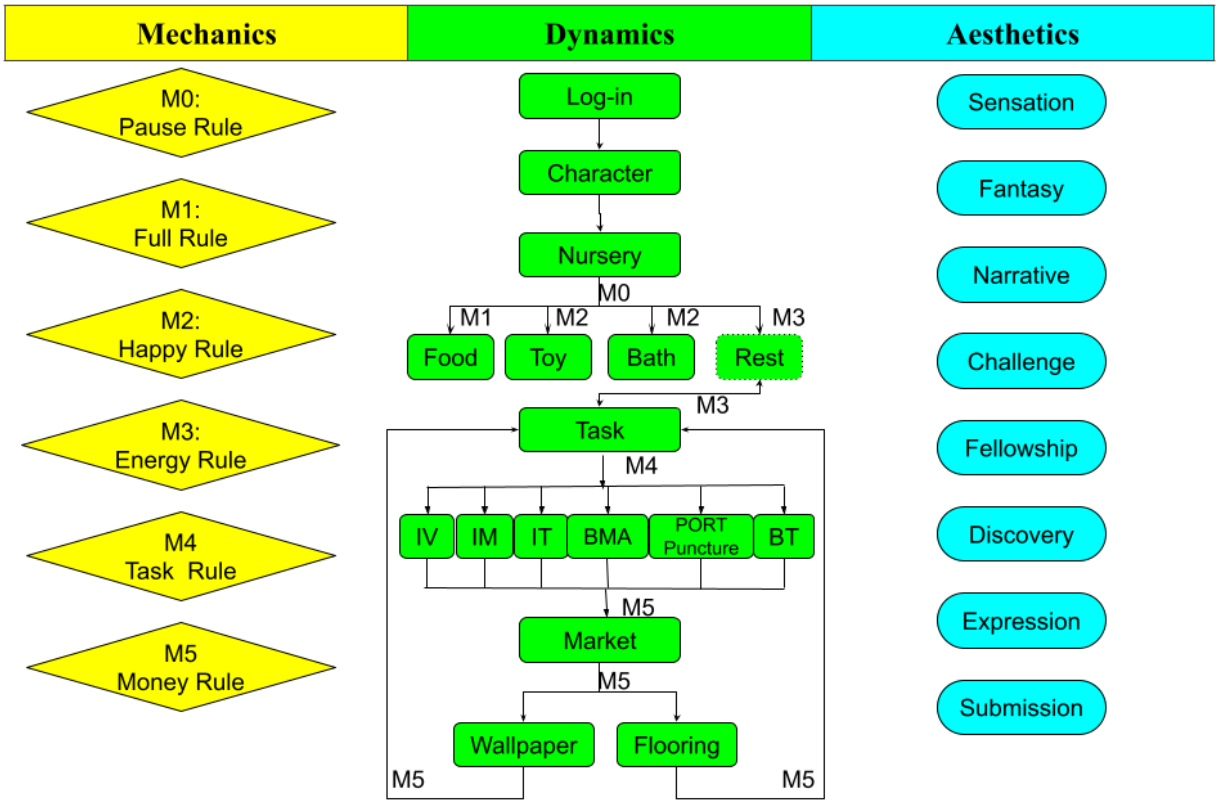
Following the MDA framework guideline (Figure 1), the TVG—named Saving the Animals' Planet—was developed using PHP and MySQL and could be played on the website.

With regard to the mechanics components, there were a total of six rules that included one or several data representations and algorithms to support the dynamics of the TVG. The "pause" rule was designed to remind the players to take breaks as needed. The "full," "happy," and "energetic" rules were designed to encourage the players to maintain good habits in their real life by eating food, playing with toys, and resting after playing the game. The "task" rule was designed to guide the players in familiarizing themselves with the procedures of the invasive therapies they would receive. The TVG made use of an encouraging voice (eg, "Good job!") and images (eg, a character setting off fireworks) as positive feedback when the player completed any task. The "money" rule was designed to

XSL•FO
RenderX

encourage the players to complete more tasks. Figure 2 shows some screenshots showing the mechanics of the game.

Figure 1. The therapeutic video game (TVG) with the MDA framework. BMA: bone marrow aspiration; BT: blood transfusion; IM: intramuscular; IT: intrathecal; IV: intravenous; MDA: Mechanics, Dynamics, and Aesthetics; NTD: New Taiwan Dollar; PORT: port-a-cath catheter system.



No.	Description of Rules
M0	❖ When playing time lasts more than 30 minutes, the TVG prompts the user and goes out of service for 10 minutes.
M1	❖ The character's "full" level (100%) is reduced by 0.0035% per second for 8 hours (eg, 100% divided by 28, 800 seconds). ❖ When the "full" level is less than 50%, the character shows the facial expression representing "hunger." ❖ The "full" level increases by 30% each time the player clicks on the food icon.
M2	❖ The character's "happy" level (100%) is reduced by 0.0012% per second for 24 hours (eg, 100% divided by 86, 400 seconds). ❖ When the "happy" level is less than 50%, the character shows the facial expression representing "unhappy." ❖ When the "happy" level is less than 30%, the character shows the facial expression representing "unhappy" and "dirty body." ❖ The "happy" level increases by 50% each time when the player clicks on the toy icon. ❖ The "happy" level increases by 40% each time when the player clicks on the bath icon. ❖ When the "happy" level is less than 30%, the player can only click on the bath icon. ❖ When the "happy" level is more than 30%, the player can click on the food and bath icons.
M3	❖ The character's "energy" level (100%) is reduced by 15% per task played. ❖ When the "energy" level is less than 30%, the character shows the facial expression representing "sleepy". ❖ When the "energy" level was less than 15%, the character is shown falling asleep. ❖ The character's "energy" level increases by 0.0056% per second for 5 hours (eg, 100% divided by18, 000 seconds).
M4	❖ The TVG randomly assigns a daily task to the player. ❖ The task is completed when the player clicks on every step sequentially. ❖ There are 6 steps for IT; 5 steps for BMA and PORT puncture; and 3 steps for IV, IM and BT.
M5	❖ The player has 20 NTDs after completing the daily task. ❖ The player has 5 NTDs after completing any of the other task. ❖ The player pays 80-100 NTDs to buy their preferred wallpaper. ❖ The player pays 80-100 NTDs to buy their preferred flooring.

XSL•FO
RenderX

Figure 2. Screenshots showing the mechanics of the therapeutic video game. "Pause" rule (left); "happy," "full," "energetic," and "money" rules (middle); and "task" (ie, complete) rule (right).



With regard to dynamics, a total of four cartoon characters—rabbit, cat, bear, and dog—were designed, and the player could choose one character as their partner. There were four statuses for each character—full, happy, energetic, and money—to show the outcome resulting from how the player nursed their partner. The statuses were dynamically changed according to the rules of the mechanics. When the character was full of energy, it could perform any of six tasks: IV injection, IM injection, IT injection, BMA, BT, or PORT puncture. Using the BMA task as an example (Figure 3), the player would have the character follow the task rule by preparing and sterilizing the needle insertion area, then covering the sterile area with hole drape; inserting the BMA needle; collecting bone marrow; and covering the wound with gauze. After completing the task, the character would be rewarded with virtual coins, with which they could buy wallpaper or flooring from the TVG market to decorate their room.

With regard to aesthetics, a total of eight emotional responses were perceived after playing the TVG: sensation, fellowship, discovery, expression, challenge, narrative, submission, and fantasy (Table 1). The sensation response was the one that provided the most aesthetics features through the anthropomorphic characters, a pleasing melody, a tender voice, and warm color through the TVG interface. The nursery feature was based on the “full,” “happy,” and “energetic” rules to encourage the players to maintain good habits in their real life. The task feature elicited the most emotional responses, as a result of the role rehearsals that took place in the medical scenarios.

XSL•FO
RenderX

Figure 3. Screenshots showing the dynamics of the bone marrow aspiration (BMA) task.

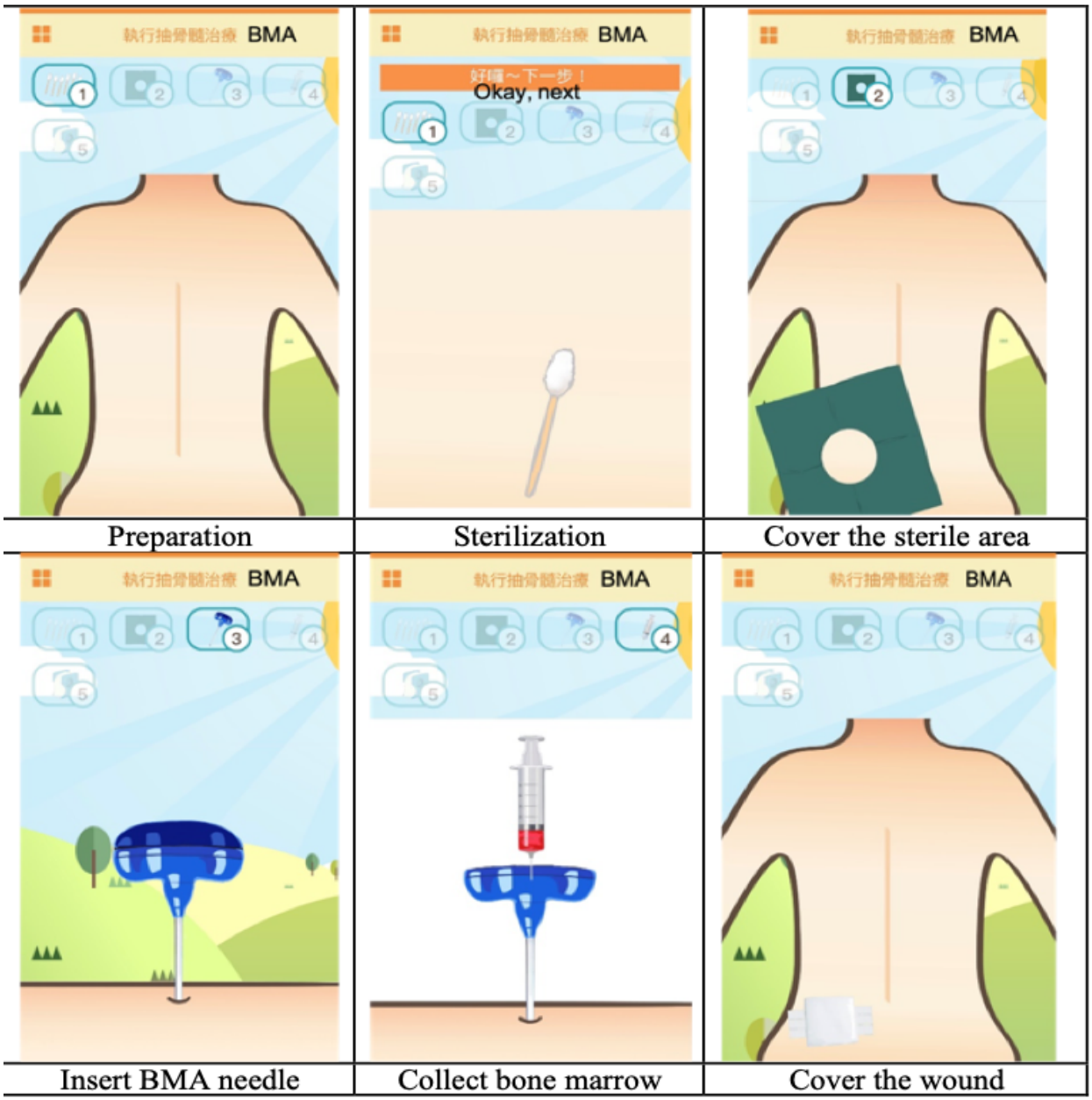


Table 1. The aesthetics of the therapeutic video game.

Feature	Aesthetics
Character	Sensation and fellowship
Nursery	Fellowship, sensation, discovery, and expression
Task	Discovery, challenge, narrative, sensation, expression, and fellowship
Market	Sensation, submission, discovery, expression, and fantasy

The Reliability of the TVG

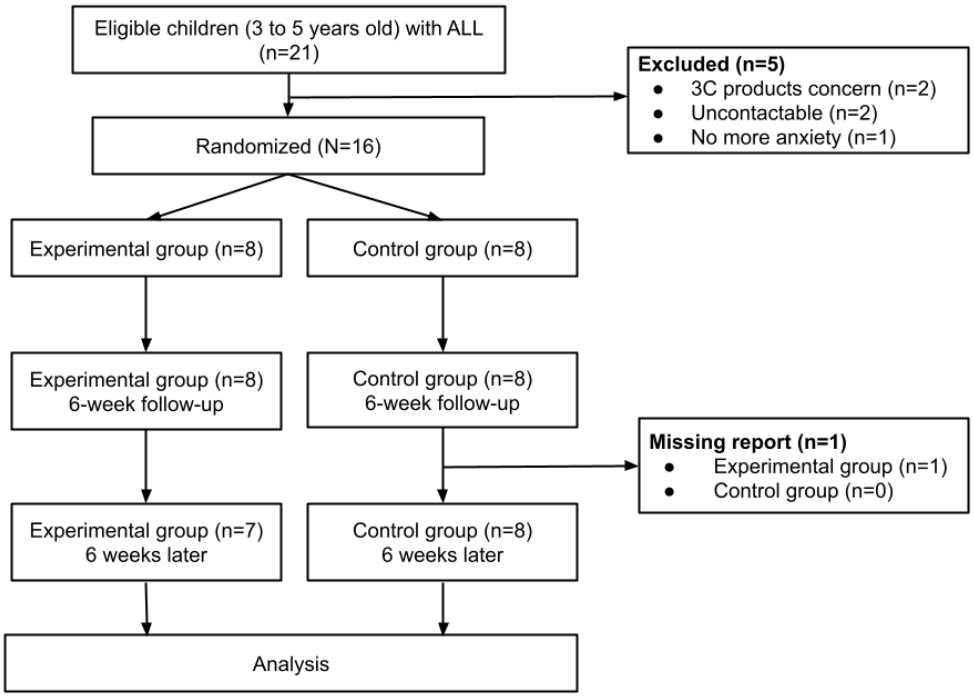
The three children's art therapists were female, had master's degrees, and were 27 to 39 years of age. Regarding the TVG, the mean usefulness score was 4.73 (SD 0.45) and the mean trust score was 4.66 (SD 0.51), which indicated that the users trusted the TVG and found it to be useful. The internal consistency (ie, Cronbach α) was .98, which indicated adequate reliability among the three therapists [20]. The narrative feedback showed that the TVG was able to (1) provide cute game elements and life simulations when completing the invasive therapy tasks, (2) have patients' characters show them encouragement when completing the tasks, and (3) provide health education knowledge to patients' family members.

XSL•FO
RenderX

Randomization and Attrition

Randomization and attrition data were organized according to the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Figure 4) [21]. Out of 21 eligible preschool-aged children with ALL, 2 family caregivers refused to participate in the trial because of concerns regarding their child using electronic products. Another 2 participants lost contact during the trial (eg, missed their clinic visiting time). An additional family caregiver said her child, who was 4 years and 10 months old, was no longer experiencing anxiety regarding receiving the invasive therapy. The remaining 16 eligible participants were enrolled and randomly allocated to the experimental group (n=8) or the control group (n=8). The recruitment rate was 76% (16/21). However, 1 participant from the experimental group dropped out because the report regarding their anxiety response was not submitted during the 6-week follow-up evaluation. A total of 7 participants remained in the experimental group, and 8 remained in the control group at the 6-week follow-up.

Figure 4. Randomized controlled trial study flowchart. 3C: computer, communication, and consumer electronics; ALL: acute lymphoblastic leukemia.



Participant Demographics

The demographics from the two groups were similar (Table 2). Most children were 5 years old (7/15, 47%), most were male (10/15, 67%), most were classified as having standard risk (11/15, 73%), and the treatment status for most of them was at the continuation stage (10/15, 67%). There were no statistically significant differences between the groups, which shows homogeneity among participants in the experimental and control groups regarding demographic characteristics.

XSL•FO
RenderX

Table 2. Demographic characteristics of the participants.

Characteristic	Experimental group (n=7), n (%)	Control group (n=8), n (%)	Total participants (N=15), n (%)	*P* value
Age in years				.95[a]
3	3 (43)	2 (25)	5 (33)	
4	0 (0)	3 (38)	3 (20)	
5	4 (57)	3 (38)	7 (47)	
Sex				>.99[b]
Male	5 (71)	5 (63)	10 (67)	
Female	2 (29)	3 (38)	5 (33)	
Risk classification				>.99[b]
Standard risk	5 (71)	6 (75)	11 (73)	
High risk	2 (29)	2 (25)	4 (27)	
Treatment status				.94[c]
Induction	2 (29)	1 (13)	3 (20)	
Consolidation	0 (0)	1 (13)	1 (7)	
Reinduction	0 (0)	1 (13)	1 (7)	
Continuation	5 (71)	5 (63)	10 (67)	

[a]This *P* value was based on the Mann-Whitney *U* test.

[b]This *P* value was based on the Fisher exact test.

[c]This *P* value was based on the chi-square test (Kendall τb).

Caregiver-Reported Invasive Therapies

According to the participants' FRS reports that were recorded by their caregivers when they received invasive therapies (Table 3), the most frequent therapy was IM injection (experimental group: 25/67, 37%; control group: 27/69, 39%; *P*=.81), and the least frequent was BT (experimental group: 2/67, 3%; control group: 0/69, 0%; *P*=.29). The total number of times that invasive therapies were received was 67 for the experimental group and 69 for the control group. The Mann-Whitney *U* test showed no statistical difference between the two groups (*P*>.05 for all).

Table 3. Caregiver-reported invasive therapies.

Invasive therapy administered	Experimental group (n=7)			Control group (n=8)			*P* value[a]
	Times administered, n (%)	Range	Mean (SD)	Times administered, n (%)	Range	Mean (SD)	
IM[b] injection (buttocks injection)	25 (37)	1-5	3.5 (1.6)	27 (39)	0-6	3.9 (1.3)	.81
PORT[c] puncture	17 (25)	0-6	2.8 (1.9)	18 (26)	1-4	2.3 (1.0)	.90
IV[d] injection	13 (19)	1-6	1.9 (1.9)	13 (19)	0-3	2.1 (0.4)	.50
IT[e] injection	6 (9)	0-2	2 (0)	8 (12)	0-2	1.1 (0.4)	.66
BMA[f]	4 (6)	0-2	2 (0)	3 (4)	0-2	1.5 (0.7)	.77
BT[g]	2 (3)	0-2	2 (0)	0 (0)	0-0	0 (0)	.29
Total	67 (100)	6-15	9.6 (3.5)	69 (100)	2-11	8.6 (2.9)	>.99

[a]This *P* value was based on the Mann-Whitney *U* test.

[b]IM: intramuscular.

[c]PORT: port-a-cath catheter system.

[d]IV: intravenous.

[e]IT: intrathecal.

[f]BMA: bone marrow aspiration.

[g]BT: blood transfusion.

XSL•FO
RenderX

Actual TVG Usage in the Experimental Group

During the 6-week follow-up evaluation in the experimental group (n=7), the average number of log-ins was 37.9 (SD 15.30, range 14-62). The average number of invasive therapy tasks completed was 58.2 (SD 59.4, range 9-179). IT injection was the most-completed task (82/408, 20.1%; Table 4). The numbers of TVG log-ins and completed tasks were not significantly correlated to treatment status (ρ=–0.138, *P*=.70).

Table 4. Actual TVG usage in the experimental group (n=7).

Item	Log-ins or completed tasks		
	n (%)	Range	Mean (SD)
TVG[a] log-ins (n=265)	265 (100)	14-62	37.9 (15.30)
TVG tasks completed (n=408)			
Total	408 (100)	9-179	58.2 (59.4)
IM[b] injection (buttocks injection)	79 (19.4)	1-28	11.3 (10.5)
PORT[c] puncture	51 (12.5)	0-26	8.5 (9.3)
IV[d] injection	66 (16.2)	0-33	11 (11.8)
IT[e] injection	82 (20.1)	0-31	13.7 (11.7)
BMA[f]	63 (15.4)	0-28	10.5 (8.9)
BT[g]	67 (16.4)	0-33	11.1 (10.9)

[a]TVG: therapeutic video game.

[b]IM: intramuscular.

[c]PORT: port-a-cath catheter system.

[d]IV: intravenous.

[e]IT: intrathecal.

[f]BMA: bone marrow aspiration.

[g]BT: blood transfusion.

Primary Outcome: Children's Anxiety Response Report

According to the participants' FRS reports, the internal consistency (ie, Cronbach α) of FRS scores was .52, which indicated acceptable reliability [20]. Table 5 shows that the mean FRS score was 6.16 (SD 3.08) among 67 records in the experimental group, and the mean FRS score was 7.45 (SD 2.71) among 69 records in the control group. The Mann-Whitney *U* test showed a statistically significant difference (*P*=.04) between the two groups (Table 5). When receiving chemical medication therapy via IV injection, the FRS score in the experimental group (mean 3.62, SD 2.63) was significantly lower than in the control group (mean 5.85, SD 3.08; *P*=.04). The other five invasive therapies showed no difference between groups. The results regarding other responses showed that the experimental group had slightly more positive responses and fewer negative responses than the control group (Table 5).

XSL•FO
RenderX

Table 5. Children's anxiety responses according to the FRS reports.

Item	Experimental group (n=7)			Control group (n=8)			*P* value
	Records (n=67), n (%)	FRS[a] score, range	FRS score, mean (SD)	Records (n=69), n (%)	FRS score, range	FRS score, mean (SD)	
Invasive therapy							
Total	67 (100)	1-10	6.16 (3.31)	69 (100)	1-10	7.45 (2.71)	.04
IM[b] injection (buttocks injection)	25 (37)	1-10	6.44 (2.97)	27 (39)	2-10	7.22 (2.83)	.42
PORT[c] puncture	17 (25)	1-10	6.82 (3.82)	18 (26)	4-10	8.22 (2.16)	.63
IV[d] injection	13 (19)	1-10	3.62 (2.63)	13 (19)	1-10	5.85 (3.08)	.04
IT[e] injection	6 (9)	4-10	7.67 (2.58)	8 (12)	5-10	8.25 (2.12)	.64
BMA[f]	4 (6)	2-10	6.5 (3.69)	3 (4)	9-10	9.67 (0.58)	.20
BT[g]	2 (3)	7-10	8.50 (2.12)	0 (0)	N/A[h]	N/A	N/A
Other responses [i]							
Positive	5 (7)	—	—	4 (6)	—	—	—
Negative	4 (6)	—	—	13 (19)	—	—	—

[a]FRS: face rating scale.

[b]IM: intramuscular.

[c]PORT: port-a-cath catheter system.

[d]IV: intravenous.

[e]IT: intrathecal.

[f]BMA: bone marrow aspiration.

[g]BT: blood transfusion.

[h]N/A: not applicable; there were no control group records for BT therapy, so there were no FRS records for that group, and the *P* value could not be calculated.

[i]The other responses items were not evaluated using the FRS, so no values appear in the related columns.

Discussion

Principal Findings

Our team developed a TVG, which was the first game specifically designed for preschool-aged children with ALL in Taiwan in order to reduce their therapy-related anxiety during their long journey of cancer treatment (Multimedia Appendix 1). Using the MDA framework, each feature of the TVG—character, nursery, tasks, and market—was designed based on the characteristics of preschool-aged children (ie, egocentrism, anthropomorphism, thinking without logic, and lacking conservation) [17]. Next, six mechanics rules (eg, pause, happy, full, energetic, money, and task) supported the dynamics of the characters in the TVG in order to complete all of the therapy-related anxiety reduction scenarios and to achieve eight aesthetics goals: sensation, fellowship, discovery, expression, narrative, submission, challenge, and fantasy (Figure 1).

According to reliability tests among the three certified children's art therapists, the results showed high agreement regarding the usefulness of the TVG (mean score 4.73, SD 0.45) and participants' trust in the TVG (mean score 4.66, SD 0.51). Two items of feedback regarding the tasks—providing cute game elements and life simulations as well as encouragement provided by characters—were adopted within the MDA framework before the clinical validation trial. The recommendation to provide health education knowledge to patients' family members was an important finding; in the future, we should think about how to involve other family members by providing additional benefits, such as this one, to those who adopt the TVG.

During the 6-week follow-up evaluation, the average number of log-ins (mean 37.9, SD 15.30) and completed tasks (mean 58.2, SD 59.4) in the experimental group showed participants' acceptance and reliability in using TVG; these were not related to treatment status (ρ=–0.138, *P*=.70). The results indicated that the participants' caregivers (eg, parents) trusted the TVG and were willing to allow their children to use it. Meanwhile, the TVG followed the TPOG-ALL 2013 protocol and was able to facilitate all stages of treatment.

Regarding the caregiver-reported children's anxiety scale, the results showed that participants in the experimental group had a significant improvement in their FRS score (mean 6.16, SD 3.31 vs 7.45, SD 2.71; *P*=.04) and fewer negative responses (mean 4, SD 6 vs mean 13, SD 19) compared to the control group. This indicated that our TVG was able to reduce therapy-related anxiety.

XSL•FO
RenderX

Comparison With Prior Work

According to a previous systematic review, improvement in therapy-related anxiety (Table 5) in our experimental group was consistent with the results of play therapy in reducing anxiety during hospital stays in different countries [22]. It was also similar to a study that found a reduction in preoperative anxiety among children aged 5 to 11 years in Jordan [23]. This may be because the TVG was designed for the preoperational stage of cognitive development of preschool-aged children, according to Piaget [17], which indicated that the participants could use the TVG easily without logical thinking. The content of the TVG also provided simulations of therapies based on real invasive therapies during ALL treatments, and it increased participants' familiarity with these therapies by allowing them to play the TVG repeatedly.

Although the TVG was able to decrease therapy-related anxiety in children with ALL, there were still some negative concerns on the part of the family caregivers. A total of 3 family caregivers from the group of eligible participants were excluded from this trial; two-thirds of these caregivers (2/3, 67%) prevented their children from using video games. To our knowledge, there is an increasing amount of literature focused on pathological and nonpathological correlates of video game playing, with specific attention directed toward internet gaming disorder [24]. It may be the cause of the lack of trust in the TVG on the part of these family caregivers.

Limitations

During the study protocol drafting phase, beginning in 2018, we adopted stratified random sampling and proposed a sample size of 34 in each group, according to the 2015-2016 Taiwan ALL incidence rate within each age and sex strata [3]. However, in Taiwan, the total number of births (ie, 165,000) was lower than the total number of deaths (ie, 173,000) for the first time in 2020 [25]. Therefore, this may have resulted in a lower ALL incidence rate among preschool-aged children during this trial (2018-2020). The statistical power for a sample size of 15 participants was 0.23, which is not adequate according to Cohen [26] and may affect the credibility of this study's results.

Another limitation is related to the study's participant blinding procedure. The primary caregivers of the participants knew their assigned group as a result of our informed consent process. To our knowledge, the effects of not blinding the patients' caregivers, which may have caused self-report bias in the RCT study, were not clear. Therefore, the improvement in therapy-related anxiety after using our TVG was conservatively inferred.

Conclusions

The purpose of this study was to investigate therapy-related anxiety responses by preschool-aged children with ALL in Taiwan after using an innovative TVG designed by our team. According to the results, the anxiety of the TVG users decreased after receiving cancer invasive therapies as compared to nonusers of the TVG. This may be because the TVG allows users to experience simulations of these invasive procedures through character role rehearsals. When the subjects received positive feedback after completing these tasks, they may have experienced cognitive and attentional distraction from their negative perceptions (ie, therapy-related anxiety).

Implications

The results from this study led to the recommendation to include more cancer-related education in the TVG, such as medication taking, hand hygiene, and toothbrushing, to help preschool-aged children with ALL take better care of themselves. In addition, the TVG could also be applied to the other common cancers (eg, brain tumors or osteosarcomas) in children to help lessen their anxiety during difficult treatments. Health education knowledge could help family members take care of their children who are ill; this could be incorporated into the design of the next TVG.

Acknowledgments

We acknowledge all the team members who participated in this research as well as the Department of Pediatrics of the National Taiwan University Hospital. We would like to give special thanks to Kun-Liang Lan and Min-Jen Hsieh for their support. We appreciate the funding from the Ministry of Science and Technology (grant 108-2314-B-010-03).

Conflicts of Interest

None declared.

Multimedia Appendix 1
TPOG-ALL 2013 protocol. ALL: acute lymphoblastic leukemia; TPOG: Taiwan Pediatric Oncology Group.
[PDF File (Adobe PDF File), 101 KB - games_v10i3e37079_app1.pdf]

Multimedia Appendix 2
User interface draft for our therapeutic video game.
[PDF File (Adobe PDF File), 11885 KB - games_v10i3e37079_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).

XSL•FO
RenderX

[PDF File (Adobe PDF File), 1378 KB - games_v10i3e37079_app3.pdf]

References

1. Onciu M. Acute lymphoblastic leukemia. Hematol Oncol Clin North Am 2009 Aug;23(4):655-674. [doi: 10.1016/j.hoc.2009.04.009] [Medline: 19577163]
2. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med 2015 Oct 15;373(16):1541-1552. [doi: 10.1056/nejmra1400972]
3. Cancer Registry Annual Report. Taipei City, Taiwan: Health Promotion Administration, Ministry of Health and Welfare; 2019. https://www.hpa.gov.tw/269/14913/n [accessed 2022-07-27].
4. Armstrong SA, Look AT. Molecular genetics of acute lymphoblastic leukemia. J Clin Oncol 2005 Sep 10;23(26):6306-6315. [doi: 10.1200/JCO.2005.05.047] [Medline: 16155013]
5. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. Lancet 2013 Jun 01;381(9881):1943-1955 [FREE Full text] [doi: 10.1016/S0140-6736(12)62187-4] [Medline: 23523389]
6. Della Starza I, Chiaretti S, De Propris MS, Elia L, Cavalli M, De Novi LA, et al. Minimal residual disease in acute lymphoblastic leukemia: Technical and clinical advances. Front Oncol 2019;9:726 [FREE Full text] [doi: 10.3389/fonc.2019.00726] [Medline: 31448230]
7. Liang D, Yang C, Lin D, Hung I, Lin K, Chen J, et al. Long-term results of Taiwan Pediatric Oncology Group studies 1997 and 2002 for childhood acute lymphoblastic leukemia. Leukemia 2010 Feb;24(2):397-405. [doi: 10.1038/leu.2009.248] [Medline: 20016538]
8. Bryant R. Managing side effects of childhood cancer treatment. J Pediatr Nurs 2003 Apr;18(2):113-125. [doi: 10.1053/jpdn.2003.11] [Medline: 12720208]
9. Katz ER, Kellerman J, Siegel SE. Behavioral distress in children with cancer undergoing medical procedures: Developmental considerations. J Consult Clin Psychol 1980;48(3):356-365. [doi: 10.1037/0022-006x.48.3.356]
10. McGrath P, Huff N. 'What is it?': Findings on preschoolers' responses to play with medical equipment. Child Care Health Dev 2001 Sep;27(5):451-462. [doi: 10.1046/j.1365-2214.2001.00219.x] [Medline: 11531917]
11. Yogman M, Garner A, Hutchinson J, Hirsh-Pasek K, Golinkoff RM, Committee on Psychosocial Aspects of Child and Family Health, Council on Communications and Media. The power of play: A pediatric role in enhancing development in young children. Pediatrics 2018 Sep;142(3):e20182058. [doi: 10.1542/peds.2018-2058] [Medline: 30126932]
12. Birnie KA, Noel M, Chambers CT, Uman LS, Parker JA. Psychological interventions for needle-related procedural pain and distress in children and adolescents. Cochrane Database Syst Rev 2018 Oct 04;10:CD005179 [FREE Full text] [doi: 10.1002/14651858.CD005179.pub4] [Medline: 30284240]
13. Arane K, Behboudi A, Goldman RD. Virtual reality for pain and anxiety management in children. Can Fam Physician 2017 Dec;63(12):932-934 [FREE Full text] [Medline: 29237632]
14. Fazelniya Z, Najafi M, Moafi A, Talakoub S. The impact of an interactive computer game on the quality of life of children undergoing chemotherapy. Iran J Nurs Midwifery Res 2017;22(6):431. [doi: 10.4103/ijnmr.ijnmr_215_15]
15. Atzori B, Hoffman HG, Vagnoli L, Patterson DR, Alhalabi W, Messeri A, et al. Virtual reality analgesia during venipuncture in pediatric patients with onco-hematological diseases. Front Psychol 2018;9:2508 [FREE Full text] [doi: 10.3389/fpsyg.2018.02508] [Medline: 30618938]
16. Hunicke R, LeBlanc M, Zubek R. MDA: A formal approach to game design and game research. In: Proceedings of the AAAI Workshop on Challenges in Game AI. 2004 Presented at: AAAI Workshop on Challenges in Game AI; July 25-26, 2004; San Jose, CA URL: https://users.cs.northwestern.edu/~hunicke/MDA.pdf
17. Piaget J. The stages of the intellectual development of the child. Bull Menninger Clin 1962 May;26:120-128. [Medline: 14486280]
18. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
19. Randomization.com. URL: http://www.jerrydallal.com/random/randomize.htm [accessed 2022-07-27]
20. Henson RK. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. Meas Eval Couns Dev 2019 Aug 29;34(3):177-189. [doi: 10.1080/07481756.2002.12069034]
21. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
22. Godino-Iáñez MJ, Martos-Cabrera MB, Suleiman-Martos N, Gómez-Urquiza JL, Vargas-Román K, Membrive-Jiménez MJ, et al. Play therapy as an intervention in hospitalized children: A systematic review. Healthcare (Basel) 2020 Jul 29;8(3):239 [FREE Full text] [doi: 10.3390/healthcare8030239] [Medline: 32751225]
23. Dwairej DA, Obeidat HM, Aloweidi AS. Video game distraction and anesthesia mask practice reduces children's preoperative anxiety: A randomized clinical trial. J Spec Pediatr Nurs 2020 Jan;25(1):e12272. [doi: 10.1111/jspn.12272] [Medline: 31576651]
24. Carpita B, Muti D, Nardi B, Benedetti F, Cappelli A, Cremone IM, et al. Biochemical correlates of video game use: From physiology to pathology. A narrative review. Life (Basel) 2021 Jul 30;11(8):775 [FREE Full text] [doi: 10.3390/life11080775] [Medline: 34440519]

XSL•FO
RenderX

25. Fertility indicators. National Development Council. URL: https://pop-proj.ndc.gov.tw/main_en/docDetail.aspx?uid=77&pid=77&docid=10016 [accessed 2022-07-28]
26. Cohen J. The concepts of power analysis. In: Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988:1.

Abbreviations

ALL: acute lymphoblastic leukemia
BMA: bone marrow aspiration
BT: blood transfusion
CONSORT: Consolidated Standards of Reporting Trials
ETP: early T-cell precursor
FRS: face rating scale
IM: intramuscular
IRB: Institutional Review Board
IT: intrathecal
IV: intravenous
MDA: Mechanics, Dynamics, and Aesthetics
PORT: port-a-cath catheter system
RCT: randomized controlled trial
TPOG: Taiwan Pediatric Oncology Group
TVG: therapeutic video game

Edited by N Zary; submitted 06.02.22; peer-reviewed by PC Wang, T Silva; comments to author 25.03.22; revised version received 01.05.22; accepted 12.06.22; published 22.08.22.

Please cite as:
Yang DJ, Lu MY, Chen CW, Liu PC, Hou IC
Development of a Therapeutic Video Game With the MDA Framework to Decrease Anxiety in Preschool-Aged Children With Acute Lymphoblastic Leukemia: Mixed Methods Approach
JMIR Serious Games 2022;10(3):e37079
URL: https://games.jmir.org/2022/3/e37079
doi:10.2196/37079
PMID:35994340

©Dai-Jie Yang, Meng-Yao Lu, Chi-Wen Chen, Pei-Ching Liu, I-Ching Hou. Originally published in JMIR Serious Games (https://games.jmir.org), 22.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cognitive Behavioral Therapy Plus a Serious Game as a Complementary Tool for a Patient With Parkinson Disease and Impulse Control Disorder: Case Report

Teresa Mena-Moreno[1,2,3], PhD; Lucero Munguía[1,2,3], PhD; Rosario Granero[2,4], PhD; Ignacio Lucas[1,3], PhD; Almudena Sánchez-Gómez[5,6], MD, PhD; Ana Cámara[5,6], BSN, MSN; Yaroslau Compta[5,6], MD, PhD; Francesc Valldeoriola[5,6], MD, PhD; Fernando Fernandez-Aranda[1,2,3,7], PhD, FAED; Anne Sauvaget[8], MD, PhD; José M Menchón[1,3,7,9], MD, PhD; Susana Jiménez-Murcia[1,2,3,7], PhD

[1]Department of Psychiatry, Bellvitge University Hospital, Hospitalet de Llobregat, Spain

[2]Centro de Investigación Biomédica en Red-Fisiopatología de la Obesidad y la Nutrición, Instituto de Salud Carlos III, Madrid, Spain

[3]Instituto de Investigación Biomédica de Bellvitge, Hospitalet de Llobregat, Spain

[4]Department of Psychobiology and Methodology, Autonomous University of Barcelona, Barcelona, Spain

[5]Parkinson's Disease and Movement Disorders Unit, Neurology Service, Hospital Clínic, Institut D'Investigacions Biomediques August Pi i Sunyer, Institut de Neurociències Universitat de Barcelona (Maria de Maeztu Excellence Center), Barcelona, Spain

[6]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Madrid, Spain

[7]Department of Clinical Sciences, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain

[8]Movement, Interactions, Performance, University of Nantes, Nantes, France

[9]Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III, Madrid, Spain

Corresponding Author:
Susana Jiménez-Murcia, PhD
Department of Psychiatry
Bellvitge University Hospital
c/ Feixa Llarga, s/n
Hospitalet de Llobregat, 08907
Spain
Phone: 34 93 2607988
Fax: 34 93 2607658
Email: sjimenez@bellvitgehospital.cat

Abstract

Background: Impulse control disorders (ICDs) are commonly developed among patients who take dopamine agonist drugs as a treatment for Parkinson disease (PD). Gambling disorder and hypersexuality are more frequent in male patients with PD, with a prevalence over 4% in dopamine agonists users. Although impulsive-compulsive behaviors are related to antiparkinsonian medication, and even though ICD symptomatology, such as hypersexuality, often subsides when the dopaminergic dose is reduced, sometimes ICD persists in spite of drug adjustment. Consequently, a multidisciplinary approach should be considered to address these comorbidities and to explore new forms of complementary interventions, such as serious games or therapies adapted to PD.

Objective: The aim of this study is to present the case of a patient with ICD (ie, hypersexuality) triggered by dopaminergic medication for PD. A combined intervention was carried out using cognitive behavioral therapy (CBT) for ICD adapted to PD, plus an intervention using a serious game—e-Estesia—whose objective is to improve emotion regulation and impulsivity. The aim of the combination of these interventions was to reduce the harm of the disease.

Methods: After 20 CBT sessions, the patient received the e-Estesia intervention over 15 sessions. Repeated measures, before and after the combined intervention, were administered to assess emotion regulation, general psychopathology, and emotional distress and impulsivity.

Results: After the intervention with CBT techniques and e-Estesia, the patient presented fewer difficulties to regulate emotion, less emotional distress, and lower levels of impulsivity in comparison to before the treatment. Moreover, the frequency and severity of the relapses also decreased.

XSL•FO
RenderX

Conclusions: The combined intervention—CBT and a serious game—showed positive results in terms of treatment outcomes.

(JMIR Serious Games 2022;10(3):e33858) doi:10.2196/33858

KEYWORDS

Parkinson disease; impulse control disorder; hypersexuality; multidisciplinary approach; serious game

Introduction

Background

Parkinson disease (PD) is a neurodegenerative disorder characterized by a set of motor disturbances, such as rigidity, slowness, and tremor [1], mainly caused by the loss of over 70% of the dopaminergic neurons in the substantia nigra. PD also leads to nonmotor symptoms, such as hyposmia, sleep disorders, and constipation; cognitive impairment disorders (eg, mild cognitive impairment and dementia) [2]; emotional disorders (eg, anxiety, 60%; apathy, 60%; and depression, 35%) [3-5]; and behavioral symptoms. In this latter regard, there is growing evidence that patients with PD are at risk of developing one or more of the four major impulse control disorders (ICDs) [6]. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, by the American Psychiatric Association [7] defines ICDs as a category of behavioral disorders characterized by recurrent maladaptive disinhibited behavior despite adverse personal and relationship consequences.

Dopaminergic medications, particularly dopamine agonists, are associated with the development of ICDs and are related to impulsive-compulsive behaviors in patients with PD [6,8]. The reported prevalence is 14%, but the estimated prevalence of each subtype of ICD is quite variable, often as a consequence of the variability in diagnostic criteria and assessment [9]. Other types of impulsive-compulsive behaviors include punding and compulsive use of dopamine replacement therapy, also known as dopamine dysregulation syndrome, whose prevalence might be over 25% among the idiopathic PD population [10].

The main risk factors for impulse control behavior in patients with PD are young age, the use of dopamine agonists, male gender, drug abuse, depression, smoking, genetic factors, impulsive personality, and family history of ICDs [9,11-20]. In patients with PD, the most frequent forms of ICD include gambling disorder (GD), which is also considered a behavioral addiction, but in any case a disorder related to impulsivity; compulsive shopping; hypersexuality; binge eating disorder, considered as an eating disorder in terms of diagnostics, but highly related to impulsivity as well; and other psychiatric complications (eg, depressive symptoms or sleep disorders) [9,21]. Regarding ICD subtypes, female patients tend to suffer more from compulsive shopping, while hypersexuality is more frequent among male patients [16,22], especially in those with an early onset of PD [16]. Hypersexuality is defined as compulsive sexual behavior, consisting of excessive and distressing sexual thoughts that persist for more than 1 month and interfere with social activity and daily routine [15,16]. It includes excessive preoccupation with sexual thoughts, desire for frequent genital stimulation, internet pornography, promiscuity, and telephone sex, among other behaviors [15,23]. It has been stated that its prevalence is over 4% in users of dopamine agonists and almost 2% in patients with PD who do not take dopamine agonists [22].

ICDs are characterized by a failure to resist a temptation, urge, or impulse that may harm oneself or others. Patients compulsively pursue certain reward-based activities and make poorly informed decisions without foreseeing the potential personal and interpersonal consequences that arise from repetitive participation in these activities. Impulsivity is a fundamental factor in ICD [24,25]. It can be understood as a trait consisting of a stable personality characteristic, which is defined as the tendency to perform behaviors without premeditation and having premature responses to stimuli that often produce adverse consequences. The model by Whiteside and Lynam [26] is one of the most accepted theoretical approaches for defining impulsivity as a multifactorial construct. It was originally formed by the following four dimensions, after which a fifth was added [27]: (1) lack of premeditation, involving acting without thinking; (2) lack of perseverance, representing the tendency to not finish tasks; (3) sensation seeking, encompassing behavior tendencies of trying new and exciting activities or sensations; and (4) positive and negative urgency, including the tendency to act rashly in response to extreme negative or positive emotions. This model has been validated across several age samples, such as children, adolescents, and young adults [28], and was the basis for development of the UPPS-P (Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency) scale [27,29].

On the other hand, the concept of emotion regulation (ER) is closely related to impulsive behaviors and addictive disorders [30-33]. Different studies have found that many addictive behaviors increased during stressful times, such as smoking and unhealthy eating [34-37] or the use and abuse of alcohol to regulate positive and negative moods [38]. Other research has found that impulsive decision-making may be an attempt to change a negative emotional state [39]. This relationship between impulsive behaviors and emotional state has been studied in previous neuroscience research, indicating that both the prefrontal cortex and the amygdala play key roles in ER [40,41], as well as in impulsive behaviors, reasoning, risk-taking, and decision-making [42-46].

Beyond its relationship with impulsivity, ER is the ability to experience and modulate emotions [47]. ER allows people to access functional resources in stressful situations and allows for the use of appropriate coping strategies [48]. Some of the strategies used to regulate emotions are adaptive (eg, reappraisal), whereas others are maladaptive (eg, suppression) [47,49]. Gratz and Roemer [50] developed the Difficulties in Emotion Regulation Scale (DERS), based on their own theoretical framework that was divided into six dimensions: nonacceptance, goals or directed behaviors, impulse control,

XSL•FO
RenderX

emotional awareness, regulation strategies, and emotional clarity. This scale has been amply validated in different age groups [51-54] and different ethnic populations [51,52,55].

Treatment Approaches

The treatment of any of the aforementioned ICDs in PD represents a challenge, due to the fact that not only is the quality of life of patients decreased, but also that of their families or caregivers [56].

Even though impulsive-compulsive behaviors are related to antiparkinsonian medication and ICD symptomatology often subsides when the dopaminergic dose is reduced [57], not all patients will decrease their dosage. Consequently, a multidisciplinary approach should be considered to address these comorbidities [58].

Jiménez-Murcia et al [59] published a protocol for the adaptation of cognitive behavioral therapy (CBT) to treat GD in patients suffering from PD. This therapy consists of 16 sessions and includes specific interventions in PD, such as psychoeducation about specific PD conditions (eg, medication checks and possible side effects) and PD's relationship with GD, as well as strategies to cope and accept a chronic illness.

Another approach that might be explored is the use of new technologies, considering that they have been shown to be a coadjuvant tool in the rehabilitation of patients with PD [60] and to help in the treatment of ICDs [61,62]. Therefore, the use of new technologies to tackle impulsive-compulsive behaviors in PD is likely to have positive responses. Finally, there are several surgical techniques indicated to reduce the side effects of antiparkinsonian medication [63]. However, the selection of the candidates must be done carefully. Neuropsychological and neuropsychiatric factors also constitute important considerations in predicting the operative outcomes [64], but the risks and benefits of the intervention must be taken into account.

A serious game is understood as "a mental contest, played with a computer [and today, with other electronic devices] in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives" [65]. In the context of mental health, serious games can be used to complement traditional psychotherapy, providing additional tools for therapists, especially in those patients that exhibit resistance to treatment and might be receptive to new technologies [66,67]. Serious games are becoming widespread resources in personalized medicine approaches, which require that treatment be tailored to the specific needs of patients [68].

A number of studies have reported promising results regarding the efficacy of serious games in treating mental health issues [69]. In particular, individuals experiencing ICDs seem to respond well to interventions aided by serious games [61,62]. In these cases, the gameplay and narrative structure are designed to favor the ER of participants, training them to delay their reactions to external and internal stimuli that usually trigger responses in them.

There is growing interest in using serious games in rehabilitation programs for neurological diseases, although to date, the literature is inconclusive and provides contradictory findings about the efficacy of game-based rehabilitation programs for PD [70]. Some studies indicate that patients with PD can benefit from serious games, as they aim at improving their physical health condition, such as their motor skills, via motion-based video games [60,71] or as they aim at helping in their rehabilitation [72]. Serious games are known to increase the rehabilitation dose by augmenting the intensity and repetition of exercises because of their motivational properties and to improve the follow-up of rehabilitation by the patient [73]. However, games conceived to tackle patients' mental health, in particular their PD-related ICDs, are also essential. Some authors emphasize the importance of the game design and personalization in order to secure better adherence and a more positive treatment outcome [74].

In addition, new therapy using computer-based ER training or serious video games has emerged [75-81].Tailor-made games that use biosignals to enhance ER have been tested under laboratory conditions. One study reported that individuals increased their emotional awareness and improved their decision-making by refining reward processing through the game's biofeedback mechanism [82]. Biofeedback can be defined as a technique in which individuals learn to adapt their behaviors based on physiological signals (eg, heart rate variability [HRV]). Thanks to the use of biofeedback, it is possible to connect the emotional reaction to the media display, which helps individuals learn how to regulate their emotions and visually rewards them when doing so appropriately [83]. A great advantage of serious games is that they allow real situations to be recreated in virtual scenarios, while generating a series of cognitive, emotional, and behavioral responses in individuals. This allows for the training of specific skills in a motivating and entertaining way that can be more difficult to achieve with traditional therapies. Self-training via serious games with biofeedback sensors has been shown to reduce general impulsive behaviors and arousal as well as to enhance self-control [61,76]. A serious game called PlayMancer presents these advantages and has shown positive results from training users in relaxation, self-awareness, and self-regulation techniques [79]. Regarding biosignals, HRV responds to autonomic flexibility; therefore, an increased HRV correlates with greater emotional control [84]. Different studies have shown the results of the use of breathing management and biofeedback with measures of HRV to reduce anxiety; reduce psychological stress, as fear of losing control; and make the use of coping strategies more habitual as a form of adaptive ER [85-87].

The aim of this study is to present the case of a patient that suffers from an ICD—hypersexuality—that is triggered by dopaminergic medication for PD. As far as we know, this is the first case study of a patient with PD with a comorbid ICD who was treated with a serious game as a complementary tool of CBT.

XSL•FO
RenderX

Methods

Case Description

The patient is a man in his early 30s who was treated at the Gambling Disorder Unit of the Department of Psychiatry at Bellvitge University Hospital, Barcelona, Spain. This is a university hospital belonging to the public health care system and is certified as a highly specialized care center for the treatment of GD and other behavioral addictions. The patient was referred from the Department of Neurology at the Hospital Clinic of Barcelona, Spain.

The patient lives with his parents, both early retirees: the father worked in the audiovisual sector, and the mother worked in a banking institution. The patient has no siblings. He studied vocational training for the position of Higher Technician in Audiovisuals and Shows. He is in a heterosexual relationship of 5 years in length. Regarding recreational activities, he used to go to the cinema and participate in sports (eg, running, football, and the gym) before the PD diagnosis. In general terms, the patient had healthy habits and did not engage in tobacco or alcohol consumption.

Clinical History

At the age of 30 years, the patient went to the outpatient clinic of the Movement Disorders Unit of the Hospital Clinic of Barcelona, having experienced a few months of predominant bradykinesia and rigidity with a mild rest tremor in his left limbs. After clinical assessment, he was diagnosed with early-onset Parkinson disease in October 2016. He started medical treatment, specifically pramipexol, with good response but with rapid tolerance, leading to a dose increase of up to 3.15 mg per day. Due to high functional requirements, levodopa/carbidopa (Sinemet Plus) was added soon after, reaching doses of 100 mg/25 mg of levodopa/carbidopa four times per day. In December 2017, after 1 year of PD evolution, the patient already presented motor fluctuations and dyskinesias as well as hypersexuality, which led to the withdrawal of the dopamine agonist and the introduction of quetiapine. Intestinal infusion of levodopa gel was eventually administered to try and deliver continuous stimulation as a means of reducing dopaminergic stimulation. Due to the lack of control of his ICD, he was referred to our unit. The patient was evaluated by a clinical psychologist using a clinical interview and the Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease; he was then diagnosed with a nonspecified ICD (ie, hypersexuality) in the context of PD.

The patient also reported visiting websites with pornographic content with the purpose of having telephone sex, as well as frequent visits to prostitutes. The frequency of the dysfunctional behavior was daily during the periods of disease activity and only decreased when the parents exercised strict external control. The patient used to run away from home, if the door was left unlocked, and disappear from his parents' sight when they went for a walk. He even used to jump out of the window of his home—it was a mezzanine—if his parents were not careful enough. In other words, the impact and interference of this behavior on his life and his family members' lives were severe.

Description of the Intervention

After the assessment, the patient started to receive psychological attention. To date, he has received 20 individual sessions of CBT adapted for cases of PD [59]. Different aspects were incorporated into the patient's treatment in order to treat his hypersexuality problem, taking into consideration his PD. A mood and mental state scan was carried out during the first and second treatment sessions. In these sessions, the patient received psychoeducation about hypersexuality and the therapeutic process and regarding specific PD conditions (eg, medication checks and possible side effects). In sessions 3 and 4, the objectives were to promote gradual abstinence (ie, stimulus control) and explain the importance of the role of cotherapists (ie, his parents). In sessions 5 and 6, difficulties and obstacles related to the treatment were explored. In addition, social skills training and productive time management (ie, planning of hobbies, role playing, and behavioral experiments) were accomplished. In sessions 7 and 8, alternative behaviors to prevent relapse were generated, and the clinical psychologist explained the relationship between cognitions, emotions, and behaviors to the patient. Moreover, specific risk situations and the development of other impulse control behaviors were analyzed, and a medical evaluation was performed in collaboration with a neurologist. From session 9 to 16, the patient performed cognitive restructuring using daily cognitive self-monitoring. Moreover, stimulus control was minimally reduced throughout these sessions, due to the enormous self-control difficulties presented by the patient. Vulnerability factors and similar aspects of thinking were discussed as well. In these sessions, the patient was trained in positive communication styles and new risk behaviors were explored. In the final therapy sessions, several objectives were established: to prevent relapse and conduct future risk factor analysis, to discuss future difficulties and the relevance of maintaining full abstinence, and to analyze the presence of other impulse control behaviors and learn how to cope with somatic chronic disease. Medical evaluation was carried out in collaboration with a neurologist.

The patient attended all the treatment sessions. He was collaborative and fully participated in the sessions. His parents acted as cotherapists throughout the treatment process. Cotherapist duties included learning about hypersexuality and PD, handling risk situations, and helping the patient with his compliance with the treatment guidelines. Furthermore, his parents directly collaborated in some aspects of the treatment, such as stimulus control, and in supporting the patient to find alternative healthy behaviors, such as new distractions and hobbies.

Although a significant improvement was observed in terms of harm reduction with CBT (ie, fewer behaviors related to seeking sexual stimulation and fewer intense and risky consequences), the relapses were still frequent: the maximum time of abstinence from the behavior was 2 weeks. He also showed high levels of impulsivity and deficits of ER, so a complementary treatment was implemented: intervention with a serious game. Thus, the patient received adjunctive treatment with a serious game called e-Estesia. This treatment aimed at improving ER and reducing impulsivity levels through the training of diaphragmatic

XSL•FO
RenderX

respiration, which might help reduce physiological activation in stressful situations. e-Estesia is a serious game that integrates the biofeedback mechanism, through heart rate and HRV biosignals, and the diaphragmatic breathing technique to induce a state of well-being in the patient. This combination is possible because cardiac fluctuations are conditioned by the respiratory processes of inspiration and expiration. Heart rate increases during inhalation of air and decreases with exhalation, which is known as respiratory sinus arrhythmia [88,89].

The e-Estesia intervention consisted of 15 sessions, each 10 minutes in length. Some questionnaires were administered on the day of the first session: the DERS; the Emotion Regulation Questionnaire (ERQ); the Symptom Checklist-90-Revised (SCL-90-R), which measures general psychopathology; and the UPPS-P scale, which measures impulsivity. These questionnaires were administered at the end of the 15 sessions.

The first session lasted 30 minutes. The patient was taught how to use the device (ie, an Android tablet), how to place the chest band with the biosensor, and how to use the serious game. He was also taught how to breathe (ie, deeply and slowly) in order to interact with the app. The benefits of this relaxation technique were also explained. The following 14 sessions with e-Estesia took place at the patient's home, without the supervision of any professional. The patient used the device at home once a day, following the instructions given in the first session. He usually played the serious game around noon, a time that coincided with the on-period of the pharmacological medication.

The Serious Game e-Estesia

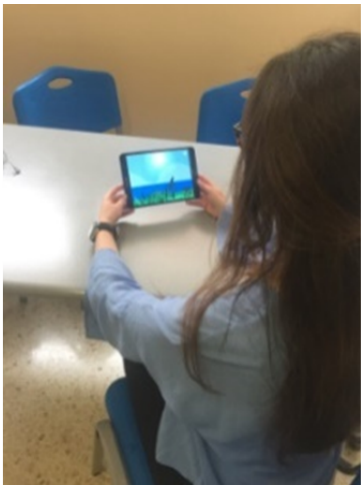
e-Estesia is a serious game that was developed by our research team at Bellvitge University Hospital. It was inspired by PlayMancer, a previous serious game that emerged during a European project at the same hospital, whose application was very successful [61,76,79,90-92]. The PlayMancer app was tested in a significant sample of patients with diagnoses of GD and eating disorders. This evaluation led to the conclusion that a future serious game should be developed using a similar strategy, but with a simpler configuration. It was also apparent to the research team that the app had to be usable at any time and in any situation, allowing the patients to bring the tablet home to self-administer the intervention as they wish. More information about the description of e-Estesia and the usability results can be found in the paper by Mena-Moreno et al [93].

The serious game e-Estesia is an app that can be used on mobile devices. It consists of a virtual marine landscape that contains a series of animations, from a radiant sun to a tropical storm. The patient must visualize the images while breathing deeply and slowly; the diaphragmatic breathing technique would have been previously taught. A physiological signal that is registered by the app is HRV, which is detected by a biosensor placed at the thoracic level (Figure 1) and is connected through Bluetooth. It allows the patient to interact with the app through the biofeedback technique while he is using the serious game (Figure 2).

Figure 1. Biosensor placed at the thoracic level to record heart rate variability.



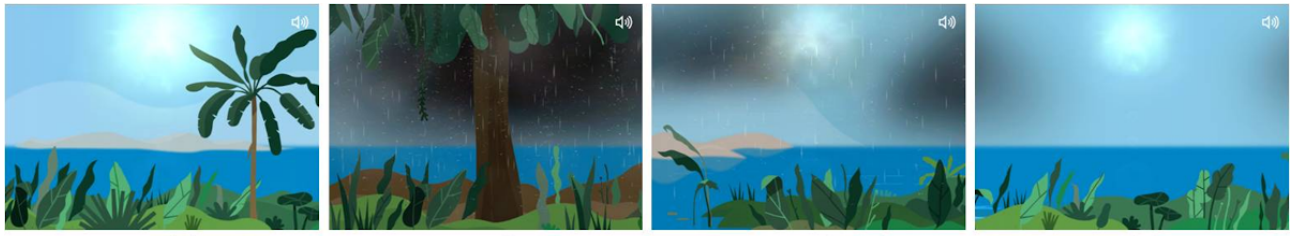
Figure 2. Use of the serious game app e-Estesia.



XSL•FO
RenderX

Each session with e-Estesia lasts 10 minutes and is divided into three periods of time. During the first and last 3 minutes, the patient should try to relax while visualizing the landscape and breathe deeply. During the middle 4 minutes, gray clouds and rain appear on the screen, simulating an intense storm; if the patient remains relaxed (ie, low activation level) and breathes deeply (ie, diaphragmatic breathing), these clouds will dissipate, which is the reward element (ie, reinforcement of the behavior to be implemented). However, if the patient stops breathing deeply and the situation agitates him or her, the storm will increase in intensity (ie, a signal that the behavior is not appropriate), the rain will be heavier, and the sound of falling water will increase. These 4 minutes allow patients to be trained to face stressful or emotionally negative situations using breathing as a complementary therapeutic strategy. These techniques are the ones learned during their standard treatment (Figure 3).

Figure 3. Animation sequence of the serous game app e-Estesia.



Minute 1-3. Initial phase. Baseline calm weather

Minute 4-7. Intermediate phase. Frustration-inducing weather

Minute 4-7. Intermediate phase. Weather regulation through emotion biofeedback

Minute 8-10. Final phase. Return to baseline calm weather

The patient plays an active role throughout the session with e-Estesia. He or she must manage their breathing (ie, breathing deeply and slowly), and the app reports feedback through the animations that appear on the tablet screen. The patient is in continuous interaction with the serious game, receiving information about whether or not he or she is performing the task adequately, reinforcing learning, while at the same time modifying his or her behavior and physiological responses to achieve the marked objective.

Ethics Approval

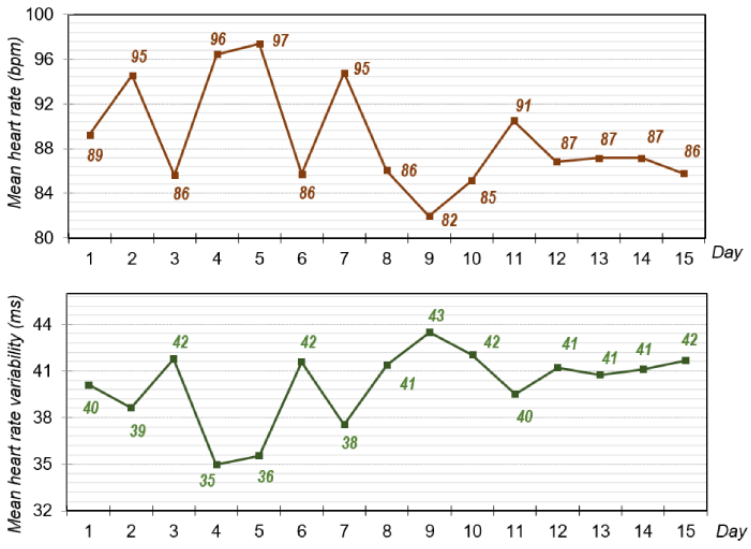
This study was conducted according to the guidelines of the Declaration of Helsinki of 1975 (revised in 2013) and was approved by the Research Ethics Committee of the Bellvitge University Hospital (reference No. PR208/22). Signed informed consent was obtained from the patient involved in the study. Since the patient was of legal age, it was not necessary for his parents to sign the consent form, although they were also informed verbally about the study by the group's principal investigator. Both the patient and his parents accepted, and the patient himself signed the informed consent document.

Results

Physiological Activity Registered in e-Estesia

Figure 4 shows the results of the heart rate and HRV measurements during 15 days of using e-Estesia. These physiological measures show two clearly different periods throughout the treatment with this system: the first half versus the second half of the treatment period. The first period included the first 7 days of practice, while the second period comprised the last 8 days. When comparing the two periods, we found that the mean heart rate decreased from the first period (92 bpm) to the second period (86.38 bpm). Regarding HRV, we also found changes between these two periods. During the first half of the treatment period (mean HRV 38.86 ms), the patient displayed lower HRV than in the second period (mean HRV 41.38 ms).

Figure 4. Mean heart rate and heart rate variability over 15 days of using the serious game app e-Estesia.



XSL•FO
RenderX

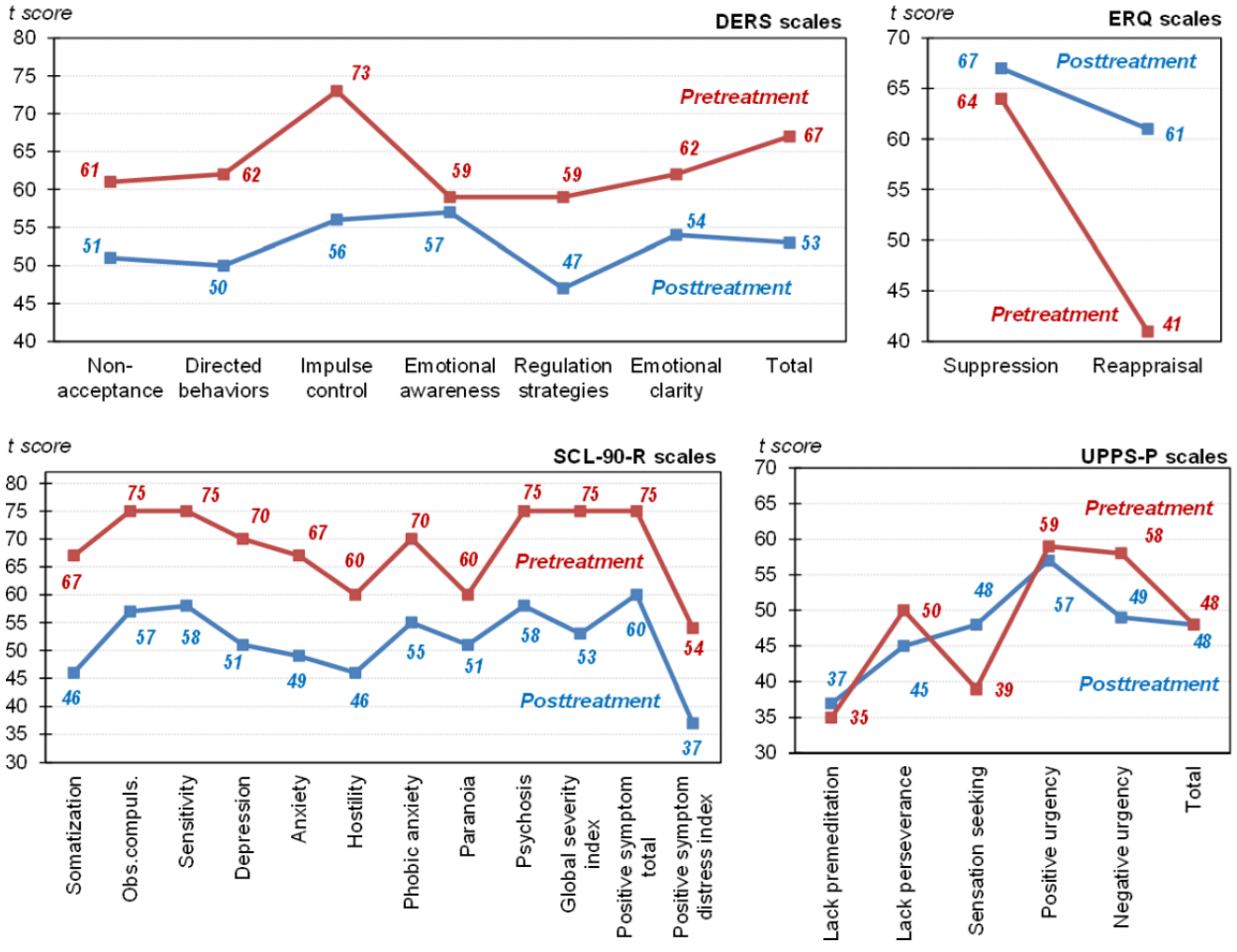
Psychopathological and Clinical Indices

Figure 5 shows the pre- to posttreatment changes in the psychometric scales of the study measuring the dysregulation of emotional state, psychopathology, and impulsivity levels. Decreasing standardized *t* scores from pre- to posttreatment were observed for the DERS: the highest change was observed for impulse control (17 points, from *t*=73 to *t*=56) and the lowest change was observed for emotional awareness (2 points, from *t*=59 to *t*=57). All the standardized *t* scores for the DERS at posttreatment were in the normal range (*t*<60) [94]. Increasing standardized *t* scores were observed from pre- to posttreatment for both strategies measured using the ERQ: reappraisal (20 points, from *t*=41 to *t*=61) and suppression (3 points, from *t*=64 to *t*=67).

For the SCL-90-R scales, decreasing standardized *t* scores from pre- to posttreatment were also observed: the highest change was for global severity index (22 points, from *t*=75 to *t*=53) and the lowest change was for paranoia (19 points, from *t*=60 to *t*=41). All the standardized *t* scores for the SCL-90-R scales at posttreatment were in the normal range (*t*<63) [95].

Regarding the UPPS-P scales, decreasing standardized *t* scores from pre- to posttreatment were also observed for lack of perseverance (5 points, from *t*=50 to *t*=45), positive urgency (2 points, from *t*=59 to *t*=57), and negative urgency (2 points, from *t*=50 to *t*=48). No pre- to posttreatment change was observed for the total score, while lack of premeditation minimally increased (2 points, from *t*=35 to *t*=37) and sensation seeking increased (9 points, from *t*=39 to *t*=48).

Figure 5. Pre- and posttest changes in the psychometric measures. DERS: Difficulties in Emotion Regulation Scale; ERQ: Emotion Regulation Questionnaire; GSI: gait symmetry index; Obs. compuls.: obsessive compulsion; premeditat.: premeditation; SCL-90-R: Symptom Checklist-90-Revised; UPPS-P: Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency.



Discussion

Principal Findings

As far as we know, this is the first case study of a patient with PD with a comorbid ICD treated with a serious game as a complementary tool to CBT. e-Estesia was employed with partial but clinically significant success, with amelioration of the symptoms related to ER and impulsivity. A significant improvement in harm reduction had already been observed with CBT (ie, fewer behaviors related to seeking sexual stimulation, with less intensity and fewer risky consequences).

After the intervention with e-Estesia, as expected, the patient presented a change in the collected psychophysiological measures (ie, heart rate and HRV). Regarding psychophysiological variables, a decrease in heart rate response was seen during the last sessions in comparison with the first ones. Regarding cardiac activity, HRV was found to be higher at the end of the e-Estesia treatment in comparison with the first

XSL•FO
RenderX

part of treatment. HRV is a measure that responds to autonomic flexibility and, therefore, an increased HRV correlates with greater emotional control [84]. Previous studies have shown that cardiac activity can be modulated by respiratory training using biofeedback applications [96-98]. Diaphragmatic breathing increases the vagal tone, activating the relaxation response of the parasympathetic nervous system [99]. Hence, the diminished heart rate response after the intervention could point toward an attenuated emotional response due to the treatment with e-Estesia. Similarly, the increase in HRV could be related to the skills learned by the patient when using e-Estesia, which probably improved ER processes. Both variables support the idea of a more adaptive affective response, probably related to better ER processes. Therefore, the treatment with e-Estesia may have produced a diminished emotional response to negative events as well as better affect regulation skills.

The results from the psychophysiological measures correspond to the changes at the clinical level collected by the psychometric measures. After the intervention with e-Estesia, an improvement was observed in general psychopathology, such as better ER and lower scores in some of the impulsivity scales.

Decreased standardized *t* scores for the DERS and the higher use of reappraisal strategies as observed by the ERQ indicated a better ER after the intervention with e-Estesia [49,50]. Reappraisal is a strategy that results in positive emotional and physical responses to emotion-eliciting stimuli, whereas various forms of suppression and avoidance have been implicated in psychopathology [47,49].

After the intervention with the serious game, the patient showed fewer difficulties in ER [50,94]. The impulse control subscale showed the greatest change, reflecting fewer difficulties in controlling impulses and behaviors in the presence of negative emotions. Also, a notable difference in the directed behaviors scale was also observed after the intervention with e-Estesia, indicating fewer difficulties in adopting goal-oriented behaviors and performing a task in the presence of negative emotions [50].

According to the impulsivity measures, there was a decrease in the scores for the positive and negative urgency UPPS-P subscales, which represent emotion-related aspects of impulsivity, with the negative urgency subscale score being even higher. These personality traits are defined as the tendency to act rashly when experiencing extreme emotional states [26,27,100], for which it is possible to hypothesize that this result might be associated with the improvements in ER observed after the intervention; as it has been mentioned in the literature, important associations between impulsivity and ER have been found in addictive disorders [30-33]. Therefore, the use of e-Estesia could be of benefit in the reduction of impulsivity reactions through the training of self-ER. As previous studies have shown, improvements in ER have a positive impact on the reduction of the performance of characteristic impulsive behaviors in other ICDs [101,102].

Another important improvement was found in the lack of perseverance UPPS-P subscale. Lack of perseverance is related to poor concentration on boring or difficult tasks [26,27], and it has been found in patients with PD who have ICD [103]. This result is in concordance with the literature, indicating that the use of serious games could be of benefit to increase attention in order to develop abilities and competencies [104,105]. Therefore, the use of e-Estesia could not only improve ER or impulsivity but other neurocognitive processes as well, such as attention and learning, which seem to be disturbed in PD. However, the level of sensation seeking continues increasing despite the treatment. This result could be related to the typical course of the disease. Higher levels of this impulsivity dimension have been found in patients with PD who have ICD [103,106,107].

Clinical symptoms measured with the SCL-90-R also showed a significant overall reduction, indicating a better psychopathological state and lower distress level at the end of the treatment [108].

Due to the severity and complexity of the case, absolute abstinence was not achieved—although, it was not the goal of the therapy—but the combined intervention (ie, CBT plus the serious game) resulted in a significant improvement in harm reduction, impulsivity, ER, and mood. It also increased the awareness and self-control of the problem behavior. In addition, the impact of this behavior on the patient's life and environment was reduced.

Limitations

Although the use of e-Estesia has been effective, due to the complexity involved in dealing with this type of comorbid problem, it is difficult to attribute the success of the intervention to the serious game only, since other factors, such as the effect of the medication, may have come into play regarding the therapeutic outcome.

Conclusions

In the clinical case presented, different interventions from a multidisciplinary perspective were carried out in order to treat ICD (ie, hypersexuality) triggered by dopaminergic medication for PD. The result of the different interventions has been positive: a combination of a manualized CBT approach plus an intervention with a serious game (e-Estesia).

Due to the different symptoms developed in this type of patient, in which the lack of impulse control itself is encouraged by the treatment (ie, usually dopaminergic medications) [6,8,9,109] for the first medical condition (ie, PD), the need arises to apply a multidisciplinary perspective. Although impulsive-compulsive behaviors are related to antiparkinsonian medication, and even though ICD symptomatology often subsides when the dopaminergic dose is reduced [57], not all patients could decrease the dosage. Hence, there is a need to apply a multidisciplinary approach to address these comorbidities [58].

In this regard, the implications of this first approximation of the use of a serious game as a coadjuvant tool in the treatment of patients with PD who have ICD are important, since positive results were found regarding the reduction of impulsivity and improvements in ER. This first approximation highlights the importance of developing multidisciplinary interventions for the treatment of complex conditions, such as comorbid PD and ICD.

XSL•FO
RenderX

Acknowledgments

This research was supported by grants from the CERCA (Centres de Recerca de Catalunya) Programme/Generalitat de Catalunya for institutional support. It was additionally supported by grants from the Ministerio de Economía y Competitividad (grant PSI2015-68701-R) and from the Ministerio de Ciencia, Innovación y Universidades (grant RTI2018-101837-B-100). The research was funded by the Delegación del Gobierno para el Plan Nacional sobre Drogas (grants 2019I47 and 2021I031), Instituto de Salud Carlos III (ISCIII; grant PI17/01167), and cofunded by the FEDER (Federacion Espanola de Enfermedades Raras) funds/European Regional Development Fund, a way to build Europe. CIBERObn (Centro de Investigación Biomédica en Red-Fisiopatología de la Obesidad y la Nutrición) and CIBERSAM (Centro de Investigación Biomédica en Red de Salud Mental) are both initiatives of ISCIII. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. TMM is supported by a predoctoral grant from the Ministerio de Educación, Cultura y Deporte (grant FPU16/02087). LM is supported by a postdoctoral grant from the Mexican institution Consejo Nacional de Ciencia y Tecnología. RG is supported by the Catalan Institution for Research and Advanced Studies (ICREA-2021 Academia Program).

Authors' Contributions

SJM, FFA, and TMM contributed to the development of the study concept and design. SJM, FFA, RG, and JMM had previous experience in the design and use of serious games for the treatment of impulsive spectrum disorders, and their experience and advice were decisive in the design of the device in this study. RG performed the statistical analysis, wrote the results, and created the figures. SJM, TMM, LM, IL, and YC aided with interpretation of the data and writing of the manuscript. TMM and LM collected the data. JMM, ASG, AC, FV, and AS revised the manuscript and provided substantive comments. SJM, FFA, and JMM obtained funding.

Conflicts of Interest

FFA received consultancy honoraria from Novo Nordisk and editorial honoraria as editor-in-chief from Wiley. The rest of the authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. GBD 2016 Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's disease, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2018 Nov;17(11):939-953 [FREE Full text] [doi: 10.1016/S1474-4422(18)30295-3] [Medline: 30287051]
2. Williams-Gray CH, Mason SL, Evans JR, Foltynie T, Brayne C, Robbins TW, et al. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. J Neurol Neurosurg Psychiatry 2013 Nov;84(11):1258-1264. [doi: 10.1136/jnnp-2013-305277] [Medline: 23781007]
3. Jankovic J. Parkinson's disease: Clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008 Apr;79(4):368-376. [doi: 10.1136/jnnp.2007.131045] [Medline: 18344392]
4. Trojano L, Papagno C. Cognitive and behavioral disorders in Parkinson's disease: An update. II: Behavioral disorders. Neurol Sci 2018 Jan;39(1):53-61. [doi: 10.1007/s10072-017-3155-7] [Medline: 29038946]
5. Schapira AHV, Chaudhuri KR, Jenner P. Non-motor features of Parkinson disease. Nat Rev Neurosci 2017 Jul;18(7):435-450. [doi: 10.1038/nrn.2017.62] [Medline: 28592904]
6. Voon V, Fox SH. Medication-related impulse control and repetitive behaviors in Parkinson disease. Arch Neurol 2007 Aug;64(8):1089-1096. [doi: 10.1001/archneur.64.8.1089] [Medline: 17698698]
7. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
8. Lawrence A, Evans A, Lees A. Compulsive use of dopamine replacement therapy in Parkinson's disease: Reward systems gone awry? Lancet Neurol 2003 Oct;2(10):595-604. [doi: 10.1016/s1474-4422(03)00529-5] [Medline: 14505581]
9. Antonini A, Barone P, Bonuccelli U, Annoni K, Asgharnejad M, Stanzione P. ICARUS study: Prevalence and clinical features of impulse control disorders in Parkinson's disease. J Neurol Neurosurg Psychiatry 2017 Apr;88(4):317-324. [doi: 10.1136/jnnp-2016-315277] [Medline: 28315845]
10. Weintraub D, David AS, Evans AH, Grant JE, Stacy M. Clinical spectrum of impulse control disorders in Parkinson's disease. Mov Disord 2015 Feb;30(2):121-127. [doi: 10.1002/mds.26016] [Medline: 25370355]
11. Callesen M, Weintraub D, Damholdt M, Møller A. Impulsive and compulsive behaviors among Danish patients with Parkinson's disease: Prevalence, depression, and personality. Parkinsonism Relat Disord 2014 Jan;20(1):22-26. [doi: 10.1016/j.parkreldis.2013.09.006] [Medline: 24090948]
12. Zhang Y, He AQ, Li L, Chen W, Liu ZG. Clinical characteristics of impulse control and related disorders in Chinese Parkinson's disease patients. BMC Neurol 2017 May 18;17(1):98 [FREE Full text] [doi: 10.1186/s12883-017-0874-6] [Medline: 28521736]

XSL•FO
RenderX

13. Joutsa J, Martikainen K, Vahlberg T, Voon V, Kaasinen V. Impulse control disorders and depression in Finnish patients with Parkinson's disease. Parkinsonism Relat Disord 2012 Feb;18(2):155-160. [doi: 10.1016/j.parkreldis.2011.09.007] [Medline: 21983019]
14. Lim S, Tan Z, Ngam P, Lor T, Mohamed H, Schee J, et al. Impulsive-compulsive behaviors are common in Asian Parkinson's disease patients: Assessment using the QUIP. Parkinsonism Relat Disord 2011 Dec;17(10):761-764. [doi: 10.1016/j.parkreldis.2011.07.009] [Medline: 21839665]
15. Maloney EM, Djamshidian A, O'Sullivan SS. Phenomenology and epidemiology of impulsive-compulsive behaviours in Parkinson's disease, atypical Parkinsonian disorders and non-Parkinsonian populations. J Neurol Sci 2017 Mar 15;374:47-52. [doi: 10.1016/j.jns.2016.12.058] [Medline: 28108021]
16. Maréchal E, Denoiseux B, Thys E, Crosiers D, Pickut B, Cras P. Impulse control disorders in Parkinson's disease: An overview from neurobiology to treatment. J Neurol 2015 Jan;262(1):7-20. [doi: 10.1007/s00415-014-7361-4] [Medline: 24824224]
17. Morgante F, Fasano A, Ginevrino M, Petrucci S, Ricciardi L, Bove F, et al. Impulsive-compulsive behaviors in parkin-associated Parkinson disease. Neurology 2016 Oct 04;87(14):1436-1441 [FREE Full text] [doi: 10.1212/WNL.0000000000003177] [Medline: 27590295]
18. Sarathchandran P, Soman S, Sarma G, Krishnan S, Kishore A. Impulse control disorders and related behaviors in Indian patients with Parkinson's disease. Mov Disord 2013 Nov;28(13):1901-1902. [doi: 10.1002/mds.25557] [Medline: 23801595]
19. Valença GT, Glass PG, Negreiros NN, Duarte MB, Ventura LM, Mueller M, et al. Past smoking and current dopamine agonist use show an independent and dose-dependent association with impulse control disorders in Parkinson's disease. Parkinsonism Relat Disord 2013 Jul;19(7):698-700. [doi: 10.1016/j.parkreldis.2013.03.004] [Medline: 23611687]
20. Weintraub D, Koester J, Potenza MN, Siderowf AD, Stacy M, Voon V, et al. Impulse control disorders in Parkinson disease: A cross-sectional study of 3090 patients. Arch Neurol 2010 May;67(5):589-595. [doi: 10.1001/archneurol.2010.65] [Medline: 20457959]
21. Weintraub D, Siderowf AD, Potenza MN, Goveas J, Morales KH, Duda JE, et al. Association of dopamine agonist use with impulse control disorders in Parkinson disease. Arch Neurol 2006 Jul;63(7):969-973 [FREE Full text] [doi: 10.1001/archneur.63.7.969] [Medline: 16831966]
22. Weintraub D, Claassen D. Impulse control and related disorders in Parkinson's disease. Int Rev Neurobiol 2017;133:679-717. [doi: 10.1016/bs.irn.2017.04.006] [Medline: 28802938]
23. O'Sullivan SS, Evans AH, Quinn NP, Lawrence AD, Lees AJ. Reckless generosity in Parkinson's disease. Mov Disord 2010 Jan 30;25(2):221-223. [doi: 10.1002/mds.22687] [Medline: 20077470]
24. Dell'Osso B, Altamura AC, Allen A, Marazziti D, Hollander E. Epidemiologic and clinical updates on impulse control disorders: A critical review. Eur Arch Psychiatry Clin Neurosci 2006 Dec;256(8):464-475 [FREE Full text] [doi: 10.1007/s00406-006-0668-0] [Medline: 16960655]
25. McElroy SL, Hudson JI, Pope HG, Keck PE, Aizley HG. The DSM-III-R impulse control disorders not elsewhere classified: Clinical characteristics and relationship to other psychiatric disorders. Am J Psychiatry 1992 Mar;149(3):318-327. [doi: 10.1176/ajp.149.3.318] [Medline: 1536268]
26. Whiteside SP, Lynam DR. The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. Pers Individ Dif 2001 Mar;30(4):669-689. [doi: 10.1016/S0191-8869(00)00064-7]
27. Verdejo-García A, Lozano O, Moya M, Alcázar MA, Pérez-García M. Psychometric properties of a Spanish version of the UPPS-P impulsive behavior scale: Reliability, validity and association with trait and cognitive impulsivity. J Pers Assess 2010 Jan;92(1):70-77. [doi: 10.1080/00223890903382369] [Medline: 20013458]
28. Cyders MA. Impulsivity and the sexes: Measurement and structural invariance of the UPPS-P Impulsive Behavior Scale. Assessment 2013 Feb;20(1):86-97. [doi: 10.1177/1073191111428762] [Medline: 22096214]
29. Whiteside SP, Lynam DR, Miller JD, Reynolds SK. Validation of the UPPS impulsive behaviour scale: A four ‐ factor model of impulsivity. Eur J Pers 2020 Dec 02;19(7):559-574. [doi: 10.1002/per.556]
30. Jakubczyk A, Trucco EM, Kopera M, Kobyliński P, Suszek H, Fudalej S, et al. The association between impulsivity, emotion regulation, and symptoms of alcohol use disorder. J Subst Abuse Treat 2018 Aug;91:49-56 [FREE Full text] [doi: 10.1016/j.jsat.2018.05.004] [Medline: 29910014]
31. Nicolai J, Darancó S, Moshagen M. Effects of mood state on impulsivity in pathological buying. Psychiatry Res 2016 Oct 30;244:351-356. [doi: 10.1016/j.psychres.2016.08.009] [Medline: 27521976]
32. Pivarunas B, Conner BT. Impulsivity and emotion dysregulation as predictors of food addiction. Eat Behav 2015 Dec;19:9-14. [doi: 10.1016/j.eatbeh.2015.06.007] [Medline: 26164390]
33. Weiss NH, Tull MT, Viana AG, Anestis MD, Gratz KL. Impulsive behaviors as an emotion regulation strategy: Examining associations between PTSD, emotion dysregulation, and impulsive behaviors among substance dependent inpatients. J Anxiety Disord 2012 Apr;26(3):453-458 [FREE Full text] [doi: 10.1016/j.janxdis.2012.01.007] [Medline: 22366447]
34. Abrantes AM, Strong DR, Lejuez CW, Kahler CW, Carpenter LL, Price LH, et al. The role of negative affect in risk for early lapse among low distress tolerance smokers. Addict Behav 2008 Nov;33(11):1394-1401 [FREE Full text] [doi: 10.1016/j.addbeh.2008.06.018] [Medline: 18684569]

XSL•FO
RenderX

35. Shi Y, Hooten WM, Warner DO. Effects of smoking cessation on pain in older adults. Nicotine Tob Res 2011 Oct;13(10):919-925 [FREE Full text] [doi: 10.1093/ntr/ntr097] [Medline: 21571690]
36. Greeno CG, Wing RR. Stress-induced eating. Psychol Bull 1994 May;115(3):444-464. [doi: 10.1037/0033-2909.115.3.444] [Medline: 8016287]
37. Magid V, Colder CR, Stroud LR, Nichter M, Nichter M, TERN Members. Negative affect, stress, and smoking in college students: Unique associations independent of alcohol and marijuana use. Addict Behav 2009 Nov;34(11):973-975 [FREE Full text] [doi: 10.1016/j.addbeh.2009.05.007] [Medline: 19523773]
38. Cooper ML, Frone MR, Russell M, Mudar P. Drinking to regulate positive and negative emotions: A motivational model of alcohol use. J Pers Soc Psychol 1995 Nov;69(5):990-1005. [doi: 10.1037//0022-3514.69.5.990] [Medline: 7473043]
39. Tice DM, Bratslavsky E, Baumeister RF. Emotional distress regulation takes precedence over impulse control: If you feel bad, do it!. J Pers Soc Psychol 2001 Jan;80(1):53-67. [Medline: 11195891]
40. Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci 2005 May;9(5):242-249. [doi: 10.1016/j.tics.2005.03.010] [Medline: 15866151]
41. Ray RD, Zald DH. Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. Neurosci Biobehav Rev 2012 Jan;36(1):479-501 [FREE Full text] [doi: 10.1016/j.neubiorev.2011.08.005] [Medline: 21889953]
42. Kim S, Lee D. Prefrontal cortex and impulsive decision making. Biol Psychiatry 2011 Jun 15;69(12):1140-1146 [FREE Full text] [doi: 10.1016/j.biopsych.2010.07.005] [Medline: 20728878]
43. Torregrossa MM, Quinn JJ, Taylor JR. Impulsivity, compulsivity, and habit: The role of orbitofrontal cortex revisited. Biol Psychiatry 2008 Feb 01;63(3):253-255 [FREE Full text] [doi: 10.1016/j.biopsych.2007.11.014] [Medline: 18194683]
44. Zeeb FD, Winstanley CA. Lesions of the basolateral amygdala and orbitofrontal cortex differentially affect acquisition and performance of a rodent gambling task. J Neurosci 2011 Feb 09;31(6):2197-2204 [FREE Full text] [doi: 10.1523/JNEUROSCI.5597-10.2011] [Medline: 21307256]
45. Spinella M. Neurobehavioral correlates of impulsivity: Evidence of prefrontal involvement. Int J Neurosci 2004 Jan;114(1):95-104. [doi: 10.1080/00207450490249347] [Medline: 14660071]
46. Krawczyk DC, Michelle McClelland M, Donovan CM. A hierarchy for relational reasoning in the prefrontal cortex. Cortex 2011 May;47(5):588-597. [doi: 10.1016/j.cortex.2010.04.008] [Medline: 20537619]
47. Gross JJ. The emerging field of emotion regulation: An integrative review. Rev Gen Psychol 1998 Sep 01;2(3):271-299. [doi: 10.1037/1089-2680.2.3.271]
48. Greenberg LS, Elliott R, Pos A. Special topic: Emotion-focused therapy. Eur Psychother 2007;7(1):87-93.
49. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: A meta-analytic review. Clin Psychol Rev 2010 Mar;30(2):217-237. [doi: 10.1016/j.cpr.2009.11.004] [Medline: 20015584]
50. Gratz K, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2008 Oct 17;30(4):315. [doi: 10.1007/s10862-008-9102-4]
51. Gómez-Simón I, Penelo E, de la Osa N. Factor structure and measurement invariance of the Difficulties Emotion Regulation Scale (DERS) in Spanish adolescents. Psicothema 2014;26(3):401-408. [doi: 10.7334/psicothema2013.324] [Medline: 25069562]
52. Marín Tejeda M, Robles García R, González–Forteza C, Andrade Palos P. [Propiedades psicométricas de la escala "Dificultades en la Regulación Emocional" en español (DERS–E) para adolescentes mexicanos] [Article in Spanish]. Salud Ment 2012;35(6):521-526 [FREE Full text]
53. Staples AM, Mohlman J. Psychometric properties of the GAD-Q-IV and DERS in older, community-dwelling GAD patients and controls. J Anxiety Disord 2012 Apr;26(3):385-392. [doi: 10.1016/j.janxdis.2012.01.005] [Medline: 22306131]
54. Weinberg A, Klonsky ED. Measurement of emotion dysregulation in adolescents. Psychol Assess 2009 Dec;21(4):616-621. [doi: 10.1037/a0016669] [Medline: 19947794]
55. Guzmán-González M, Trabucco C, Urzúa M A, Garrido L, Leiva J. Validez y confiabilidad de la versión adaptada al español de la escala de Dificultades de Regulación Emocional (DERS-E) en población chilena [Article in Spanish]. Ter Psicol 2014 Apr;32(1):19-29 [FREE Full text] [doi: 10.4067/s0718-48082014000100002]
56. Giladi N, Weitzman N, Schreiber S, Shabtai H, Peretz C. New onset heightened interest or drive for gambling, shopping, eating or sexual activity in patients with Parkinson's disease: The role of dopamine agonist treatment and age at motor symptoms onset. J Psychopharmacol 2007 Jul;21(5):501-506. [doi: 10.1177/0269881106073109] [Medline: 17446202]
57. Ferreri F, Agbokou C, Gauthier S. Recognition and management of neuropsychiatric complications in Parkinson's disease. CMAJ 2006 Dec 05;175(12):1545-1552 [FREE Full text] [doi: 10.1503/cmaj.060542] [Medline: 17146092]
58. Thanvi BR, Munshi SK, Vijaykumar N, Lo TCN. Neuropsychiatric non-motor aspects of Parkinson's disease. Postgrad Med J 2003 Oct;79(936):561-565 [FREE Full text] [doi: 10.1136/pmj.79.936.561] [Medline: 14612597]
59. Jiménez-Murcia S, Bove F, Israel M, Steiger H, Fernández-Aranda F, Alvarez-Moya E, et al. Cognitive-behavioral therapy for pathological gambling in Parkinson's disease: A pilot controlled study. Eur Addict Res 2012;18(6):265-274 [FREE Full text] [doi: 10.1159/000337442] [Medline: 22760081]

XSL•FO
RenderX

60. Dauvergne C, Bégel V, Gény C, Puyjarinet F, Laffont I, Dalla Bella S. Home-based training of rhythmic skills with a serious game in Parkinson's disease: Usability and acceptability. Ann Phys Rehabil Med 2018 Nov;61(6):380-385 [FREE Full text] [doi: 10.1016/j.rehab.2018.08.002] [Medline: 30193992]
61. Tárrega S, Castro-Carreras L, Fernández-Aranda F, Granero R, Giner-Bartolomé C, Aymamí N, et al. A serious videogame as an additional therapy tool for training emotional regulation and impulsivity control in severe gambling disorder. Front Psychol 2015;6:1721 [FREE Full text] [doi: 10.3389/fpsyg.2015.01721] [Medline: 26617550]
62. Fagundo AB, Santamaría JJ, Forcano L, Giner-Bartolomé C, Jiménez-Murcia S, Sánchez I, et al. Video game therapy for emotional regulation and impulsivity control in a series of treated cases with bulimia nervosa. Eur Eat Disord Rev 2013 Nov;21(6):493-499. [doi: 10.1002/erv.2259] [Medline: 24092598]
63. Lahtinen MJ, Haapaniemi TH, Kauppinen MT, Salokorpi N, Heikkinen ER, Katisko JP. A comparison of indirect and direct targeted STN DBS in the treatment of Parkinson's disease-surgical method and clinical outcome over 15-year timespan. Acta Neurochir (Wien) 2020 May;162(5):1067-1076 [FREE Full text] [doi: 10.1007/s00701-020-04269-x] [Medline: 32103343]
64. Wagle Shukla A, Okun MS. Surgical treatment of Parkinson's disease: Patients, targets, devices, and approaches. Neurotherapeutics 2014 Jan;11(1):47-59 [FREE Full text] [doi: 10.1007/s13311-013-0235-0] [Medline: 24198187]
65. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
66. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: Current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
67. Fovet T, Micoulaud-Franchi J, Vaiva G, Thomas P, Jardri R, Amad A. [Serious game as a therapeutic tool in psychiatry: A systematic review] [Article in French]. Encephale 2016 Oct;42(5):463-469. [doi: 10.1016/j.encep.2016.02.008] [Medline: 27017317]
68. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform 2012;177:85-96. [Medline: 22942036]
69. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: Are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
70. Barry G, Galna B, Rochester L. The role of exergaming in Parkinson's disease rehabilitation: A systematic review of the evidence. J Neuroeng Rehabil 2014 Mar 07;11:33 [FREE Full text] [doi: 10.1186/1743-0003-11-33] [Medline: 24602325]
71. Assad O, Hermann R, Lilla D, Mellies B, Meyer R, Shevach L, et al. Motion-based games for Parkinson's disease patients. In: Proceedings of the 10th International Conference on Entertainment Computing. 2011 Presented at: 10th International Conference on Entertainment Computing; October 5-8, 2011; Vancouver, BC p. 47-58. [doi: 10.1007/978-3-642-24500-8_6]
72. Paraskevopoulos IT, Tsekleves E, Craig C, Whyatt C, Cosmas J. Design guidelines for developing customised serious games for Parkinson's Disease rehabilitation using bespoke game sensors. Entertain Comput 2014 Dec;5(4):413-424. [doi: 10.1016/j.entcom.2014.10.006]
73. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
74. Dias S, Konstantinidis E, Diniz J, Bamidis P, Charisis V, Hadjidimitriou S, et al. Serious games as a means for holistically supporting Parkinson's Disease patients: The i-PROGNOSIS personalized game suite framework. In: Proceedings of the 9th International Conference on Virtual Worlds and Games for Serious Applications. 2017 Presented at: 9th International Conference on Virtual Worlds and Games for Serious Applications; September 6-8, 2017; Athens, Greece p. 237-244. [doi: 10.1109/vs-games.2017.8056607]
75. Jon Hobbs L, Yan Z. Cracking the walnut: Using a computer game to impact cognition, emotion, and behavior of highly aggressive fifth grade students. Comput Human Behav 2008 Mar;24(2):421-438. [doi: 10.1016/j.chb.2007.01.031]
76. Jiménez-Murcia S, Fernández-Aranda F, Kalapanidas E, Konstantas D, Ganchev T, Kocsis O, et al. Playmancer project: A serious videogame as an additional therapy tool for eating and impulse control disorders. Stud Health Technol Inform 2009;144:163-166. [Medline: 19592756]
77. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
78. Ducharme P, Wharff E, Hutchinson E, Kahn J, Logan G, Gonzalez-Heydrich J. Videogame assisted emotional regulation training: An ACT with RAGE-control case illustration. Clin Soc Work J 2011 Sep 27;40(1):75-84. [doi: 10.1007/s10615-011-0363-0]
79. Fernández-Aranda F, Jiménez-Murcia S, Santamaría JJ, Gunnard K, Soto A, Kalapanidas E, et al. Video games as a complementary therapy tool in mental disorders: PlayMancer, a European multicentre study. J Ment Health 2012 Aug;21(4):364-374 [FREE Full text] [doi: 10.3109/09638237.2012.664302] [Medline: 22548300]
80. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: A systematic review. Games Health J 2018 Apr;7(2):85-99. [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]

XSL•FO
RenderX

81. Bettis A, Burke T, Nesi J, Liu R. Digital technologies for emotion-regulation assessment and intervention: A conceptual review. Clin Psychol Sci 2022 Jan;10(1):3-26 [FREE Full text] [doi: 10.1177/21677026211011982] [Medline: 35174006]
82. Astor P, Adam M, Jerčić P, Schaaff K, Weinhardt C. Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. J Manag Inf Syst 2014 Dec 09;30(3):247-278. [doi: 10.2753/mis0742-1222300309]
83. Jerčić P, Sundstedt V. Practicing emotion-regulation through biofeedback on the decision-making performance in the context of serious games: A systematic review. Entertain Comput 2019 Mar;29:75-86. [doi: 10.1016/j.entcom.2019.01.001]
84. Appelhans B, Luecken L. Heart rate variability as an index of regulated emotional responding. Rev Gen Psychol 2006 Sep 01;10(3):229-240. [doi: 10.1037/1089-2680.10.3.229]
85. Lehrer P, Vaschillo E, Vaschillo B, Lu S, Eckberg D, Edelberg R, et al. Heart rate variability biofeedback increases baroreflex gain and peak expiratory flow. Psychosom Med 2003;65(5):796-805. [doi: 10.1097/01.psy.0000089200.81962.19] [Medline: 14508023]
86. Gross MJ, Shearer DA, Bringer JD, Hall R, Cook CJ, Kilduff LP. Abbreviated resonant frequency training to augment heart rate variability and enhance on-demand emotional regulation in elite sport support staff. Appl Psychophysiol Biofeedback 2016 Sep;41(3):263-274. [doi: 10.1007/s10484-015-9330-9] [Medline: 26782084]
87. Gruzelier JH, Thompson T, Redding E, Brandt R, Steffert T. Application of alpha/theta neurofeedback and heart rate variability training to young contemporary dancers: State anxiety and creativity. Int J Psychophysiol 2014 Jul;93(1):105-111. [doi: 10.1016/j.ijpsycho.2013.05.004] [Medline: 23684733]
88. Ataee P, Belingard L, Dumont G, Noubari H, Boyce W. Autonomic-cardiorespiratory regulation: A physiology-based mathematical model. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 28-September 1, 2012; San Diego, CA p. 3805-3808. [doi: 10.1109/embc.2012.6346796]
89. Yasuma F, Hayano J. Respiratory sinus arrhythmia: Why does the heartbeat synchronize with respiratory rhythm? Chest 2004 Feb;125(2):683-690. [doi: 10.1378/chest.125.2.683] [Medline: 14769752]
90. Claes L, Jiménez-Murcia S, Santamaría JJ, Moussa MB, Sánchez I, Forcano L, et al. The facial and subjective emotional reaction in response to a video game designed to train emotional regulation (Playmancer). Eur Eat Disord Rev 2012 Nov;20(6):484-489. [doi: 10.1002/erv.2212] [Medline: 23097370]
91. Giner-Bartolomé C, Fagundo AB, Sánchez I, Jiménez-Murcia S, Santamaría JJ, Ladouceur R, et al. Can an intervention based on a serious videogame prior to cognitive behavioral therapy be helpful in bulimia nervosa? A clinical case study. Front Psychol 2015;6:982 [FREE Full text] [doi: 10.3389/fpsyg.2015.00982] [Medline: 26236261]
92. Guerrero-Vaca D, Granero R, Fernández-Aranda F, Mestre-Bach G, Martín-Romera V, Mallorquí-Bagué N, et al. Explicit and implicit emotional expression in gambling disorder measured by a serious game: A pilot study. J Gambl Stud 2021 Jun;37(2):467-481. [doi: 10.1007/s10899-020-09945-2] [Medline: 32253655]
93. Mena-Moreno T, Fernández-Aranda F, Granero R, Munguía L, Steward T, López-González H, et al. A serious game to improve emotion regulation in treatment-seeking individuals with gambling disorder: A usability study. Front Psychol 2021;12:621953 [FREE Full text] [doi: 10.3389/fpsyg.2021.621953] [Medline: 33746839]
94. Wolz I, Agüera Z, Granero R, Jiménez-Murcia S, Gratz K, Menchón JM, et al. Emotion regulation in disordered eating: Psychometric properties of the Difficulties in Emotion Regulation Scale among Spanish adults and its interrelations with personality and clinical severity. Front Psychol 2015;6:907 [FREE Full text] [doi: 10.3389/fpsyg.2015.00907] [Medline: 26175710]
95. Gonzalez De Rivera JL, Derogatis LR, De las Cuevas C, Gracia Marco R, Rodríguez-Pulido F, Henry-Benitez M, et al. The Spanish Version of the SCL-90-R. Normative Data in the General Population. Towson, MD: Clinical Psychometric Research; 1989.
96. Chittaro L, Sioni R. Evaluating mobile apps for breathing training: The effectiveness of visualization. Comput Human Behav 2014 Nov;40:56-63. [doi: 10.1016/j.chb.2014.07.049]
97. Dillon A, Kelly M, Robertson IH, Robertson DA. Smartphone applications utilizing biofeedback can aid stress reduction. Front Psychol 2016;7:832 [FREE Full text] [doi: 10.3389/fpsyg.2016.00832] [Medline: 27378963]
98. Lin I. Effects of a cardiorespiratory synchronization training mobile application on heart rate variability and electroencephalography in healthy adults. Int J Psychophysiol 2018 Dec;134:168-177. [doi: 10.1016/j.ijpsycho.2018.09.005] [Medline: 30243751]
99. Lehrer P, Gevirtz R. Heart rate variability biofeedback: How and why does it work? Front Psychol 2014;5:756 [FREE Full text] [doi: 10.3389/fpsyg.2014.00756] [Medline: 25101026]
100. Cyders M, Smith GT. Emotion-based dispositions to rash action: Positive and negative urgency. Psychol Bull 2008 Nov;134(6):807-828 [FREE Full text] [doi: 10.1037/a0013341] [Medline: 18954158]
101. Mallorquí-Bagué N, Mena-Moreno T, Granero R, Vintró-Alcaraz C, Sánchez-González J, Fernández-Aranda F, et al. Suicidal ideation and history of suicide attempts in treatment-seeking patients with gambling disorder: The role of emotion dysregulation and high trait impulsivity. J Behav Addict 2018 Dec 01;7(4):1112-1121 [FREE Full text] [doi: 10.1556/2006.7.2018.132] [Medline: 30580545]

XSL•FO
RenderX

102. Axelrod SR, Perepletchikova F, Holtzman K, Sinha R. Emotion regulation and substance use frequency in women with substance dependence and borderline personality disorder receiving dialectical behavior therapy. Am J Drug Alcohol Abuse 2011 Jan;37(1):37-42 [FREE Full text] [doi: 10.3109/00952990.2010.535582] [Medline: 21091162]
103. Hlavatá P, Linhartová P, Šumec R, Filip P, Světlák M, Baláž M, et al. Behavioral and neuroanatomical account of impulsivity in Parkinson's disease. Front Neurol 2019;10:1338 [FREE Full text] [doi: 10.3389/fneur.2019.01338] [Medline: 31998210]
104. Drummond D, Hadchouel A, Tesnière A. Serious games for health: Three steps forwards. Adv Simul (Lond) 2017;2:3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
105. Loh CS, Sheng Y, Ifenthaler D. Serious games analytics: Theoretical framework. In: Loh CS, Sheng Y, Ifenthaler D, editors. Serious Games Analytics. Advances in Game-Based Learning. Cham, Switzerland: Springer International Publishing; 2015:3-29.
106. Bayard S, Joly E, Ghisletta P, Rossignol A, Herades Y, Geny C, et al. A multidimensional approach to impulsivity in Parkinson's disease: Measurement and structural invariance of the UPPS Impulsive Behaviour Scale. Psychol Med 2016 Jul 27;46(14):2931-2941. [doi: 10.1017/s0033291716001586]
107. Voon V, Sohr M, Lang AE, Potenza MN, Siderowf AD, Whetteckey J, et al. Impulse control disorders in Parkinson disease: A multicenter case-control study. Ann Neurol 2011 Jun;69(6):986-996. [doi: 10.1002/ana.22356] [Medline: 21416496]
108. Derogatis LR, Unger R. Symptom Checklist-90-Revised. In: The Corsini Encyclopedia of Psychology. Hoboken, NJ: John Wiley & Sons, Inc; 2010.
109. Garcia-Ruiz PJ, Martinez Castrillo JC, Alonso-Canovas A, Herranz Barcenas A, Vela L, Sanchez Alonso P, et al. Impulse control disorder in patients with Parkinson's disease under dopamine agonist therapy: A multicentre study. J Neurol Neurosurg Psychiatry 2014 Aug;85(8):840-844. [doi: 10.1136/jnnp-2013-306787] [Medline: 24434037]

Abbreviations

CBT: cognitive behavioral therapy
CERCA: Centres de Recerca de Catalunya
CIBERObn: Centro de Investigación Biomédica en Red-Fisiopatología de la Obesidad y la Nutrición
CIBERSAM: Centro de Investigación Biomédica en Red de Salud Mental
DERS: Difficulties in Emotion Regulation Scale
ER: emotion regulation
ERQ: Emotion Regulation Questionnaire
FEDER: Federacion Espanola de Enfermedades Raras
GD: gambling disorder
HRV: heart rate variability
ICD: impulse control disorder
ISCIII: Instituto de Salud Carlos III
PD: Parkinson disease
SCL-90-R: Symptom Checklist-90-Revised
UPPS-P: Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency

Edited by N Zary; submitted 28.09.21; peer-reviewed by E Rahmani, Z Aghaei, N Chalghaf, V Verma; comments to author 25.03.22; revised version received 03.04.22; accepted 22.04.22; published 09.09.22.

Please cite as:
Mena-Moreno T, Munguía L, Granero R, Lucas I, Sánchez-Gómez A, Cámara A, Compta Y, Valldeoriola F, Fernandez-Aranda F, Sauvaget A, Menchón JM, Jiménez-Murcia S
Cognitive Behavioral Therapy Plus a Serious Game as a Complementary Tool for a Patient With Parkinson Disease and Impulse Control Disorder: Case Report
JMIR Serious Games 2022;10(3):e33858
URL: https://games.jmir.org/2022/3/e33858
doi:10.2196/33858
PMID:36083621

©Teresa Mena-Moreno, Lucero Munguía, Rosario Granero, Ignacio Lucas, Almudena Sánchez-Gómez, Ana Cámara, Yaroslau Compta, Francesc Valldeoriola, Fernando Fernandez-Aranda, Anne Sauvaget, José M Menchón, Susana Jiménez-Murcia. Originally published in JMIR Serious Games (https://games.jmir.org), 09.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Effectiveness of Serious Games on Cognitive Processing Speed Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis

Alaa Abd-alrazaq[1], PhD; Arfan Ahmed[1], PhD; Haitham Alali[2], PhD; Ahmad Mohammad Aldardour[3], MSc; Mowafa Househ[4], PhD

[1]AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

[2]Health Management Department, Faculty of Medical and Health Sciences, Liwa College of Technology, Abu Dhabi, United Arab Emirates

[3]Physical Therapy Department, Rumailah Hospital, Hamad Medical Corporation, Doha, Qatar

[4]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
PO Box 34110
Doha Al Luqta St, Ar-Rayyan
Doha, 0000
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Human cognitive processing speed is known to decline with age. Human cognitive processing speed refers to the time that an individual takes from receiving a stimulus to reacting to it. Serious games, which are video games used for training and educational purposes, have the potential to improve processing speed. Numerous systematic reviews have summarized the evidence regarding the effectiveness of serious games in improving processing speed, but they are undermined by some limitations.

Objective: This study aimed to examine the effectiveness of serious games on the cognitive processing speed of an older adult population living with cognitive impairment.

Methods: A systematic review of randomized controlled trials (RCTs) was conducted. Two search sources were used in this review: 8 electronic databases and backward and forward reference list checking. A total of 2 reviewers independently checked the eligibility of the studies, extracted data from the included studies, and appraised the risk of bias and quality of the evidence. Evidence from the included studies was synthesized using a narrative and statistical approach (ie, meta-analysis), as appropriate.

Results: Of the 548 publications identified, 16 (2.9%) RCTs eventually met all eligibility criteria. Very-low-quality evidence from 50% (8/16) and 38% (6/16) of the RCTs showed no statistically significant effect of serious games on processing speed compared with no or passive intervention groups (P=.77) and conventional exercises (P=.58), respectively. A subgroup analysis showed that both types of serious games (cognitive training games: P=.26; exergames: P=.88) were as effective as conventional exercises in improving processing speed.

Conclusions: There is no superiority of serious games over no or passive interventions and conventional exercises in improving processing speed among older adults with cognitive impairment. However, our findings remain inconclusive because of the low quality of the evidence, the small sample size in most of the included studies, and the paucity of studies included in the meta-analyses. Therefore, until more robust evidence is published, serious games should be offered or used as an adjunct to existing interventions. Further trials should be undertaken to investigate the effect of serious games that specifically target processing speed rather than cognitive abilities in general.

Trial Registration: PROSPERO CRD42022301667; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=301667

XSL•FO
RenderX

(JMIR Serious Games 2022;10(3):e36754) doi:10.2196/36754

KEYWORDS

serious games; cognitive training; exergames; processing speed; mild cognitive impairment; Alzheimer disease; dementia; systematic reviews; meta-analysis; mobile phone

Introduction

Background

By 2050, the older adult population is expected to double on a global scale [1] followed by a growing demand for high-quality support among this population. There is a variety of disabilities related to aging. However, a considerable percentage of people aged >65 years are affected by cognitive impairments. According to the Alzheimer's Association, approximately 15% to 20% of people aged >65 years are affected by mild cognitive impairment (MCI) [2]. It is a progressive, terminal brain disorder for which the cure and cause are unknown [3]. Worldwide, the number of Alzheimer disease (AD) cases is rising; estimates indicate that 44 million people currently have AD or dementia. As the Baby Boomer generation passes the age of 65 years, the number of people experiencing AD is expected to reach 76 million by 2030 [4,5]. The consequences of these cognitive disorders include poor judgment, decreased activity, and memory loss. They are difficult to manage; thus, it would be highly beneficial to develop proper and effective interventions for the older adult population.

These disorders imply a decline in various cognitive functions such as memory [6,7], executive function [8,9], attention [10], and processing speed [11]. Processing speed refers to the time that an individual takes from receiving a stimulus to reacting to it. Slowed processing speed is a typical characteristic of cognitive aging. It implies a slowing rate at which individuals perform decision-making as well as motor and perceptual tasks. Moreover, processing speed represents one of the critical predictors of performance related to cognitive tasks in older adults; it is the basis of a primary hypothesis for age-related cognitive decline [11,12]. Accordingly, aging research primarily aims to design methods for improving and maintaining cognitive functions in older adults.

Serious games are one of the interventions that have been used to improve and maintain processing speed in older adults. Serious games represent a type of video game used for training and educational purposes in different contexts [13,14]. Serious games have shown promising results in improving cognitive functions such as global cognition [15-18], processing speed [16-20], memory [16-23], executive function [16-22,24], and attention [16,17,19,20,22]. The most common types of serious games used for improving cognitive abilities are exergames (video games that require physical exercise as part of playing) and cognitive training games (video games that include cognitively stimulating activities designed to maintain or promote the users' cognitive abilities).

Research Gap and Aim

Several studies have explored the therapeutic impact of serious games on processing speed. However, the evidence from these studies is still fragmented. Aggregating the evidence through systematic reviews is important to draw conclusions about the effectiveness of serious games in improving processing speed. A total of 5 systematic reviews have pooled findings from these studies, but they are undermined by some limitations [16-20]. Specifically, these reviews (1) focused on older adults without cognitive impairment [16,18-20], (2) included quasi-experiments or pilot randomized controlled trials (RCTs) [17-19], (3) did not assess the quality of the meta-analyzed evidence [17,18,20], (4) only focused on a specific type of serious game such as cognitive training games [16,17,20] and exergames [18], or (5) did not compare the effect of serious games with a specific comparator (eg, no intervention, conventional exercises, or conventional cognitive activities) [16-18,20]. Accordingly, the aim of this review was to examine the effectiveness of serious games on the cognitive processing speed of an older adult population living with cognitive impairment. To address the aforementioned limitations, we (1) focused on older adults with cognitive impairment; (2) included only RCTs; (3) appraised the quality of the meta-analyzed evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach; (4) included all types of serious games; and (5) compared the effect of serious games with that of a specific comparator. In this review, effectiveness refers to the degree to which serious games are successful in improving cognitive processing speed.

Methods

To conduct this systematic review, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [25]. The protocol for this review is registered at PROSPERO (ID: CRD42022301667).

Search Strategy

Search Sources

The first author searched the following databases on November 10, 2021: MEDLINE (via Ovid), PsycINFO (via Ovid), Embase (via Ovid), CINAHL (via EBSCO), IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar. Only the first 10 pages (ie, 100 hits) were considered as the databases return a large number of studies automatically ordered according to relevance [26]. Finally, backward reference list checking (ie, screening the reference lists of the included studies and relevant reviews) and forward reference list checking (ie, screening the studies that cited the included studies) were conducted.

Search Terms

The search query was developed in consultation with 2 experts in digital mental health. It included terms related to the target population (eg, cognitive disorder), target intervention (eg, serious games), and targeted study design (eg, RCTs).

XSL•FO
RenderX

Multimedia Appendix 2 summarizes the search query that was used for each of the 8 databases.

Study Eligibility Criteria

Only RCTs that evaluated the effectiveness of serious games in improving processing speed among older adults with cognitive impairment were included in this study. To be more precise, we considered studies that included serious games available on any digital platform, such as PCs, video game consoles (eg, Xbox and PlayStation), mobile phones, tablets, handheld devices, Nintendo, or any other type of digital device. The game had to be the major component of the intervention and used solely for therapeutic purposes. Studies involving serious games in combination with other interventions were included if the control group underwent the same adjacent intervention. Games that were not based on digital technology (eg, paper-and-pencil games and board games) or that were used for monitoring, screening, diagnosis, or research were excluded.

The study population was older adults (aged ≥60 years) with any type of cognitive impairment or disorder as confirmed by checking the inclusion criteria or baseline scores against defined diagnostic criteria (eg, Mini-Mental State Examination). Older adults without cognitive impairment, health care providers, and caregivers were beyond the scope of this review. No restrictions were applied regarding gender or ethnicity.

The outcome of interest in this review was cognitive processing speed. No restrictions were applied regarding the outcome measures. This review did not consider studies that focused only on cost-effectiveness, acceptance, feasibility, satisfaction, or cognitive abilities other than processing speed. The focus of this review was on postintervention data (ie, outcome data collected just after the intervention) rather than follow-up data (ie, outcome data collected a period after the intervention).

All types of RCTs (ie, parallel, cluster, crossover, and factorial) were considered in this review, whereas pilot RCTs, quasi-experiments, observational studies, and reviews were excluded. We included journal articles, conference proceedings, and dissertations, whereas abstracts, conference posters, commentaries, proposals, and editorials were excluded. This review was restricted to only studies written in the English language and published since 2010. We did not apply restrictions on the country of publication, comparator, or study settings.

Study Selection

We identified relevant studies by using the following process. First, all the retrieved studies were imported into EndNote (Clarivate Analytics) to find duplicate publications and remove them. Second, the titles and abstracts of all the retrieved studies were checked by 2 reviewers working independently (the first and second authors). Finally, both reviewers independently read the complete texts of the studies included in the previous step. Any disagreements were resolved via discussion. Steps 2 and 3 had an interrater agreement (Cohen κ) of 0.88 and 0.96, respectively.

Data Extraction

Before extracting the data, we pilot-tested the data extraction form with 2 of the included studies. Microsoft Excel was used by 2 reviewers (the first and second authors) to independently extract data from the included studies. Any disputes in the extracted data between the reviewers were resolved through discussion. The first and corresponding authors of the included studies were contacted to retrieve outcome data (eg, mean, SD, and sample size) if they were missing from the published articles. The data extraction form is provided in Multimedia Appendix 3.

Risk of Bias Appraisal

Two reviewers (the first and second authors) independently examined the risk of bias in the included studies using the Risk of Bias 2 tool [27]. This tool evaluates the risk of bias in 5 areas of RCTs: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the results [27]. The reviewers held discussions to resolve any disagreements, and the interrater agreement was 0.85.

Data Synthesis

Narrative and statistical methods were used to summarize the collected data. In our narrative synthesis, we used text and tables to describe the characteristics of the included studies (study metadata, population, interventions, comparisons, and outcome measures). The results of the experiments were compiled and classified according to the comparator: no or passive interventions, conventional exercises, and other serious games. When 2 or more studies from the same comparator submitted sufficient data (ie, mean, SD, and number of participants in each intervention group), meta-analyses were performed using Review Manager (RevMan 5.4; The Cochrane Collaboration). As the type of data for the outcome of interest (processing speed) was continuous and the instruments used to evaluate the outcome varied across the included trials, the standardized mean difference (SMD; Cohen *d*) was used to estimate the overall effect of each study. We also chose the random effects model for the analysis because of the high clinical heterogeneity among the meta-analyzed trials in terms of serious game characteristics (eg, type, duration, frequency, and period), population characteristics (eg, sample size, mean age, and health condition), and outcome measures (ie, tools and follow-up period).

To assess the degree and statistical significance of heterogeneity in the meta-analyzed studies, we calculated I^2 and a chi-square *P* value, respectively. A chi-square *P* value of ≤.05 indicated heterogeneous meta-analyzed studies [28]. When I^2 ranged from 0% to 40%, 30% to 60%, 50% to 90%, and 75% to 100%, the degree of heterogeneity was considered as insignificant, moderate, substantial, or considerable, respectively [28].

We used the GRADE approach [29] to appraise the overall quality of the evidence resulting from the meta-analyses. The GRADE approach examines the quality of evidence based on 5 domains: risk of bias, inconsistency (ie, heterogeneity), indirectness, imprecision, and publication bias [29]. The quality of the meta-analyzed evidence was independently assessed by 2 reviewers (the first and second authors). Any differences between the reviewers were resolved via discussion, and the interrater agreement of the reviewers was 0.94 [30].

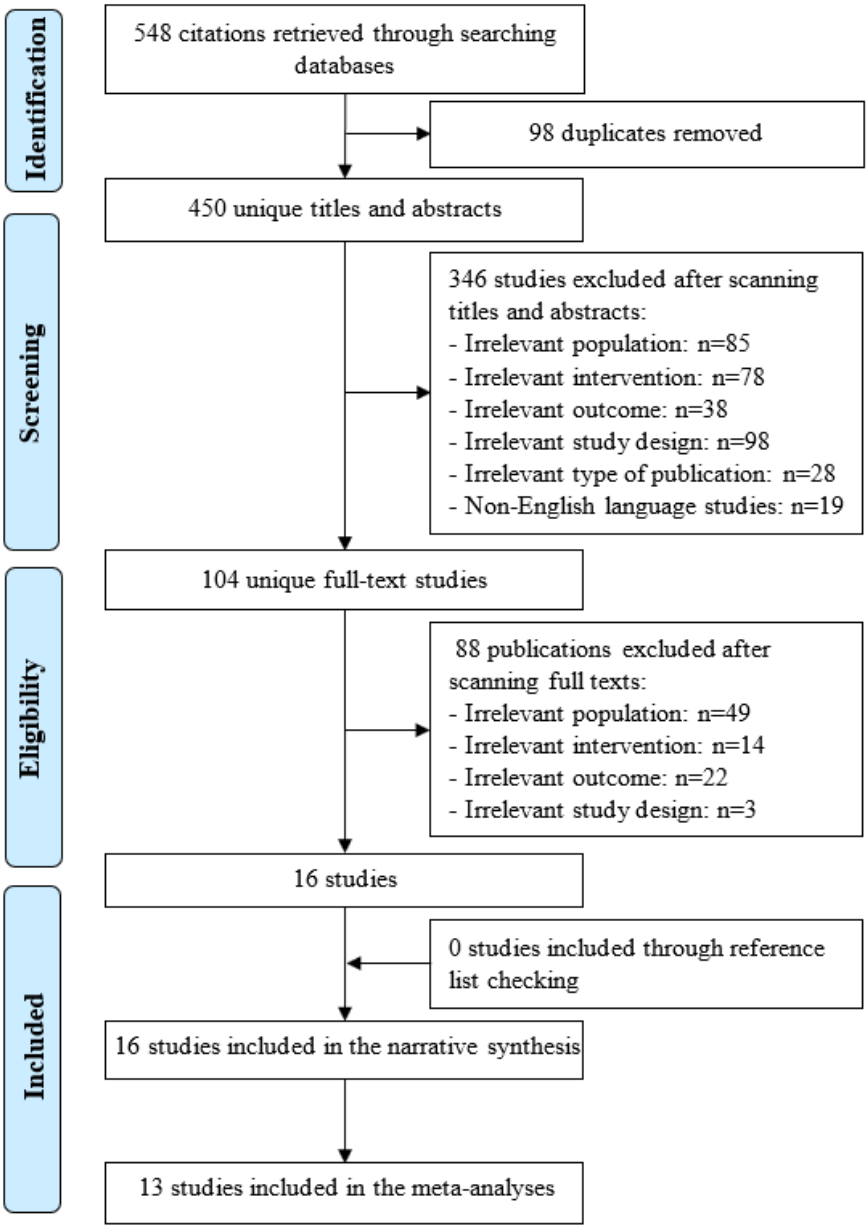
XSL•FO
RenderX

Results

Search Results

As shown in Figure 1, a total of 548 records were found after searching 8 electronic databases. Using the EndNote software, of the 548 records, 98 (17.9%) duplicates were removed. Checking the titles and abstracts of the remaining records resulted in the exclusion of 63.1% (346/548) for several reasons shown in Figure 1. Reading the full text of the remaining 104 publications resulted in the exclusion of 88 (84.6%) studies (Figure 1). This review included 16 RCTs in total [31-46]. Of these 16 RCTs, 13 (81%) were included in the meta-analyses [31-43].

Figure 1. Flowchart of the study selection process.



Characteristics of the Included Studies

The included studies were published between 2012 and 2021 and came from 11 different countries (Table 1). Except for a book chapter, all the included papers (15/16, 94%) were peer-reviewed articles. Parallel RCTs were the most common type of trial used in the included studies (14/16, 88%). The sample sizes of the included studies ranged from 20 to 195, with an average of 79.9. The average age of the participants was reported in 94% (15/16) of the studies and varied between 66 and 82.9 years, with an average of 75 years. Male participants in 94% (15/16) of the studies ranged from 23.5% to 71%, with an average of 47.5%. The mean Mini-Mental State Examination score of the participants was reported in 88% (14/16) of the studies and ranged from 18.6 to 28.1, with an average of 24.4. The most common disorder among the participants in the included studies was MCI (11/16, 69%). Participants were drawn from clinical (10/16, 62%), community (4/16, 25%), and clinical and community (2/16, 12%) settings.

XSL•FO
RenderX

Table 1. Characteristics of the studies and populations (N=16).

Study, year	Country	Publication type	RCT[a] type	Sample size	Age, mean (SD)	Male participants, n (%)	MMSE[b] score	Health condition	Setting
Finn and McDonald [31], 2015	Australia	Journal article	Parallel	31	75.6	22 (71)	28.1	MCI[c]	Clinical
Robert et al [32], 2020	France	Journal article	Parallel	46	79.4	22 (47.8)	21.4	Neurocognitive disorders	Clinical
Savulich et al [33], 2017	United Kingdom	Journal article	Parallel	42	76.1	25 (59.5)	26.7	MCI	Clinical
Valdes et al [34], 2012	United States	Journal article	Parallel	195	77.7	65 (33.3)	NR[d]	MCI	Clinical
Yang and Kwak [35], 2017	South Korea	Journal article	Parallel	20	71	14 (70)	23.1	AD[e]	Clinical
Thapa et al [36], 2020	South Korea	Journal article	Parallel	68	72.7	16 (23.5)	26.2	MCI	Clinical
Tarnanas et al [37], 2014	Greece	Book chapter	Parallel	114	70.3	44 (39)	26.4	MCI	Clinical
Fiatarone Singh et al [38], 2014	Australia	Journal article	Factorial	100	70.1	32 (32)	27	MCI	Community
Amjad et al [39], 2019	Pakistan	Journal article	Parallel	44	NR	NR	24	MCI	Clinical
Wiloth et al [40], 2017	Germany	Journal article	Parallel	99	82.9	28 (28.3)	22	Dementia	Clinical and community
Van Santen et al [41], 2020	Netherlands	Journal article	Cluster	112	79	60 (53.5)	18.6	Dementia	Clinical
Karssemeijer et al [42], 2019	Netherlands	Journal article	Parallel	115	79.9	62 (53.9)	22.4	Dementia	Clinical and community
Liao et al [43], 2021	Taiwan	Journal article	Parallel	61	81.5	20 (32.6)	22.9	MCI	Community
Flak et al [44], 2019	Norway	Journal article	Parallel	85	66	57 (66.7)	NR	MCI	Clinical
Hyer et al [45], 2016	United States	Journal article	Parallel	68	75.2	32 (47.1)	26	MCI	Community
Park and Park [46], 2017	South Korea	Journal article	Parallel	78	67.3	42 (53.8)	26.5	MCI	Community

[a]RCT: randomized controlled trial.

[b]MMSE: Mini-Mental State Examination.

[c]MCI: mild cognitive impairment.

[d]NR: not reported.

[e]AD: Alzheimer disease.

We identified 16 distinct serious games used in the studies (Table 2). More than one game was used in 6% (1/16) of the studies. The serious games in the trials were divided into 2 categories based on the treatment modality they provided: cognitive training games (12/16, 75%) and exergames (4/16, 25%). In 88% (14/16) of the studies, the games were designed with a "serious" objective from the start (designed serious games). By contrast, games in the remaining 12% (2/16) of the studies were not planned as serious games from the start but were instead used for a serious purpose (purpose-shifted games). Serious games were played under the supervision of health care providers or caregivers in most studies (11/16, 69%). The game duration in the included studies ranged from 25 to 100 minutes. The frequency of playing the games ranged from 2 to 7 times per week, but it was <4 times per week in approximately two-thirds of the studies (11/16, 69%). The duration of the interventions ranged from 4 to 25 weeks, but it was <13 weeks in three-fourths of the studies (13/16, 81%).

XSL•FO
RenderX

Table 2. Characteristics of the interventions (N=16).

Study	Serious game name	Serious game type	Serious game genre	Platform	Supervision	Duration (minutes)	Frequency (times per week)	Period (weeks)
Finn and McDonald [31]	E-Prime	Cognitive training game	Designed	PC	Supervised	NR[a]	2	4
Robert et al [32]	MeMo	Cognitive training game	Designed	PC and tablet	Unsupervised	30	4	12
Savulich et al [33]	Game Show	Cognitive training game	Designed	Tablet	Supervised	60	2	4
Valdes et al [34]	SOPT	Cognitive training game	Designed	PC	Supervised	60	2	5
Yang and Kwak [35]	Brain-Care	Cognitive training game	Designed	PC	Unsupervised	60	2	12
Thapa et al [36]	Juice Making, Crow Shooting, Love House, and Fireworks	Cognitive training game	Designed	VR[b] headset and hand controllers	Supervised	100	3	8
Tarnanas et al [37]	Virtual Reality Museum	Cognitive training game	Designed	VR headset	Supervised	90	2	21
Fiatarone Singh et al [38]	COGPACK	Cognitive training game	Designed	PC	Supervised	75	2	25
Amjad et al [39]	Body and Brain Exercises	Cognitive training game	Purpose-shifted	Xbox console and Kinect	Supervised	25 to 30	5	6
Wiloth et al [40]	Physiomat	Exergame	Designed	Balance broad and screen	Supervised	90	2	10
Van Santen et al [41]	NR	Exergame	Designed	Stationary bicycle and screen	Unsupervised	NR	5	25
Karssemeijer et al [42]	NR	Exergame	Purpose-shifted	Stationary bicycle and screen	Supervised	30 to 50	3	12
Liao et al [43]	Tano and Long-Good	Exergame	Designed	Kinect and VR headset	Supervised	60	3	12
Flak et al [44]	Cogmed	Cognitive training game	Designed	PC	Unsupervised	30 to 40	5	5
Hyer et al [45]	Cogmed	Cognitive training game	Designed	PC	Both	40	7	5 to 7
Park and Park [46]	CoTras	Cognitive training game	Designed	PC	Supervised	30	3	10

[a]NR: not reported.

[b]VR: virtual reality.

Half of the studies (8/16, 50%) delivered no or passive interventions to the comparison groups (eg, reading newspaper articles, surfing the internet, or watching a documentary program), whereas these groups received active interventions (eg, conventional exercises and other serious games) in 62% (10/16) of the studies (Table 3). A total of 12% (2/16) of the studies delivered both active and passive interventions as comparators. The active interventions had a duration ranging from 25 to 100 minutes. The active interventions were performed once to 7 times per week. The active comparator duration ranged from 5 to 25 weeks. The outcome of interest (ie, processing speed) was assessed using 11 different tools, with the Trail Making Test A being a frequently used tool in the included studies (7/16, 44%). The outcome of interest was measured after the intervention in all the included studies (16/16, 100%). In total, 44% (7/16) of the studies followed the participants after the interventions, and the follow-up period varied between 4 and 261 weeks. The number of participants who dropped out of the included studies ranged from 0 to 28.

XSL•FO
RenderX

Table 3. Characteristics of the comparators and outcomes (N=16).

Study	Comparator	Duration (minutes)	Frequency (times per week)	Period (weeks)	Outcome measures	Follow-up	Attritions, N
Finn and McDonald [31]	Control	N/A[a]	N/A	N/A	D-KEFS-NS[b]	After the intervention	7
Robert et al [32]	Control	N/A	N/A	N/A	TMT-A[c], SCWT[d], and WAIS-R-Dsy[e]	After the intervention; 12-week follow-up	NR[f]
Savulich et al [33]	Control	N/A	N/A	N/A	CANTAB-CRT[g]	After the intervention	0
Valdes et al [34]	Control	N/A	N/A	N/A	UFOV[h]	After the intervention; 52-, 104-, 156-, and 261-week follow-up	NR
Yang and Kwak [35]	Control	N/A	N/A	N/A	GnG[i] and SCWT	After the intervention	0
Thapa et al [36]	Control	N/A	N/A	N/A	TMT-A and SDST[j]	After the intervention	2
Tarnanas et al [37]	Control and conventional cognitive activities	90	2	21	SCWT	After the intervention	9
Fiatarone Singh et al [38]	Control, conventional exercises+sham cognitive training, and serious games+conventional exercises	Control: 75; conventional exercises+sham cognitive training: 100; serious games+conventional exercises: 60	2	25	SDMT[k]	After the intervention; 74-week follow-up	14
Amjad et al [39]	Conventional exercises	25 to 30	5	6	TMT-A	After the intervention	6
Wiloth et al [40]	Conventional exercises	60	2	10	TMT-A	After the intervention; 12-week follow-up	26
Van Santen et al [41]	Conventional exercises	N/A	5	25	TMT-A	Midintervention and after the intervention	28
Karssemeijer et al [42]	Conventional exercises (aerobic exercises); conventional exercises (relaxation and flexibility exercises)	30 to 50	3	12	TMT-A and SCWT	Midintervention and after the intervention; 24-week follow-up	23
Liao et al [43]	Conventional exercises	60	3	12	SCWT	After the intervention	15
Flak et al [44]	Nonadaptive serious game	30 to 40	5	5	D-KEFS-CWIT1[l] and D-KEFS-CWIT2[m]	After the intervention; 4- and 16-week follow-up	17
Hyer et al [45]	Nonadaptive serious game	40	7	5 to 7	TMT-A	After the intervention; 12-week follow-up	9
Park and Park [46]	Exergames	30	3	10	SCWT	After the intervention	0

[a]N/A: not applicable.

[b]D-KEFS-NS: Delis-Kaplan Executive Function System Test Battery-Number Sequencing.

[c]TMT-A: Trail Making Test A.

[d]SCWT: Stroop Color and Word Test.

[e]WAIS-R-Dsy: Wechsler Adult Intelligence Scale-Revised-Digit Symbol.

XSL•FO
RenderX

[f]NR: not reported.

[g]CANTAB-CRT: Cambridge Neuropsychological Test Automated Battery-Choice Reaction Time.

[h]UFOV: Useful Field of View test.

[i]GnG: go-no go.

[j]SDST: Symbol Digit Substitution Test.

[k]SDMT: Symbol Digit Modalities Test.

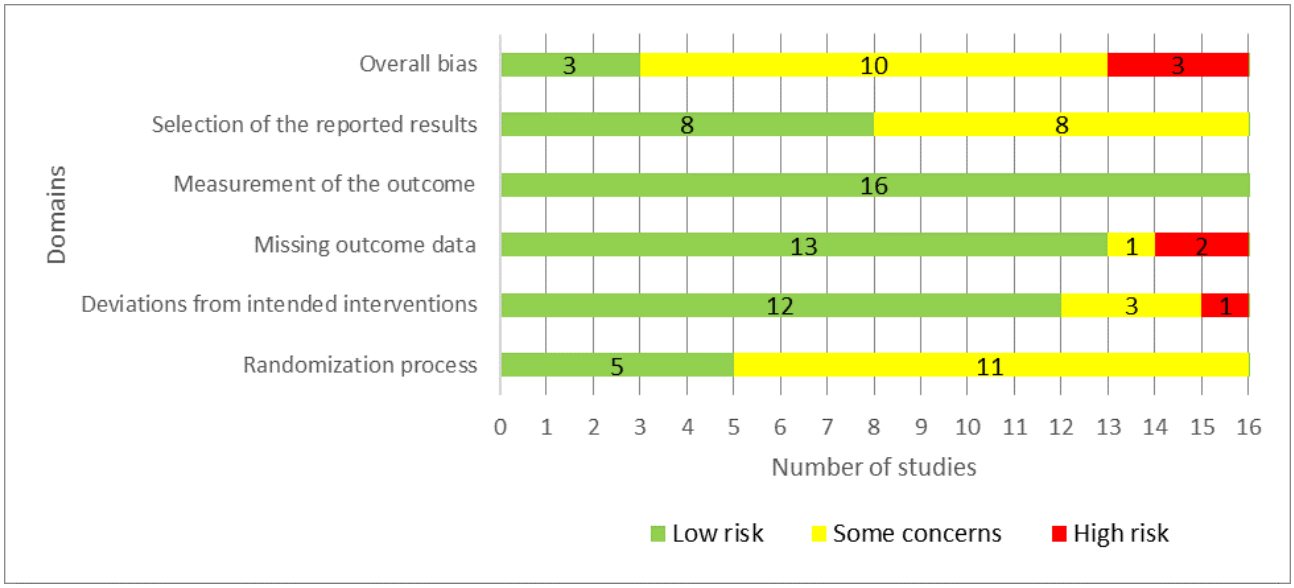
[l]D-KEFS-CWIT1: Delis-Kaplan Executive Function System Color-Word Interference Test 1.

[m]D-KEFS-CWIT2: Delis-Kaplan Executive Function System Color-Word Interference Test 2.

Results of Risk of Bias Appraisal

As shown in Figure 2, a total of 31% (5/16) of the studies were judged to have a low risk of bias in the "randomization process" domain. With regard to the "deviations from the intended interventions" domain, there was a low risk of bias in 75% (12/16) of the studies. The risk of bias because of missing outcome data was low in 81% (13/16) of the studies. All the included studies (16/16, 100%) were judged to have a low risk of bias in the "measuring the outcome" domain. In half of the included studies (8/16, 50%), the risk of bias was rated as low in the "selection of the reported results" domain. According to these judgments, only 19% (3/16) of the studies were judged to have a low risk of bias in the last domain (ie, overall bias). Reviewers' judgments about each "risk of bias" domain for each included study are presented in Multimedia Appendix 4 [31-46].

Figure 2. Review authors' judgments about each "risk of bias" domain.



Results of the Studies

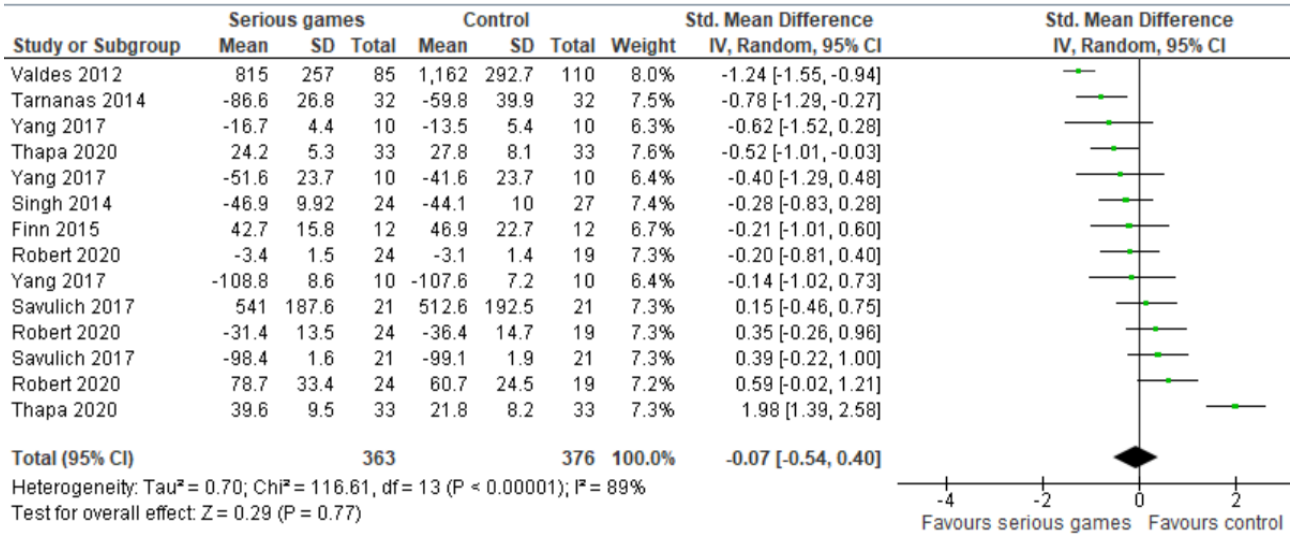
Serious Games Versus No or Passive Interventions

The effect of serious games was compared with that of control (no or passive interventions) in 50% (8/16) of the studies [31-38]. Passive interventions are those that do not have a known effect on the measured outcome, such as reading newspaper articles, surfing the internet, and watching a documentary program. In 50% (4/8) of these studies [32,33,35,36], more than one outcome measure was used to assess processing speed. Therefore, we included the results of all these measures in the meta-analysis to form 14 comparisons. As shown in Figure 3 [31-38], there was no statistically significant difference (P=.77) in processing speed between serious games and the control groups (SMD: −0.07, 95% CI −0.54 to 0.40). The statistical heterogeneity of the evidence was considerable (P<.001; I^2=89%). The quality of the evidence was very low as it was downgraded by 6 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5).

The SMDs of 2 comparisons seem to be outliers (−1.24 [34] and 1.98 [36]), although the characteristics of the studies in these comparisons were comparable with those of other studies in this meta-analysis. For this reason, we ran a sensitivity analysis to assess whether removing these outliers influenced the overall effect size and heterogeneity level. The sensitivity analysis showed that the difference in processing speed between the groups remained insignificant (P=.32), but the heterogeneity substantially decreased from 89% to 49%.

XSL•FO
RenderX

Figure 3. Forest plot of 8 studies (14 comparisons) comparing the effect of serious games on processing speed with that of control [31-38].

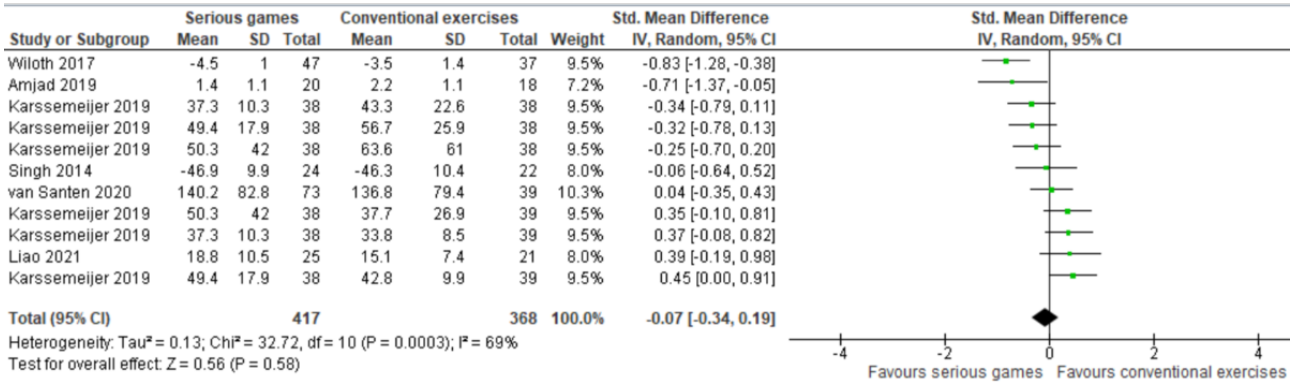


Serious Games Versus Conventional Exercises

The effect of serious games was compared with that of conventional exercises in 38% (6/16) of the studies [38-43]. Of these 6 studies, 1 (17%) compared serious games with 2 different conventional exercises (aerobic exercises and relaxation and flexibility exercises) and measured processing speed using 2 different tools [42]. Therefore, we included the results of all these comparisons and measures in the meta-analysis to form 11 comparisons (Figure 4 [38-43]). The meta-analysis showed no statistically significant difference (P=.58) in processing speed between the serious game and conventional exercise groups (SMD: −0.07, 95% CI −0.34 to 0.19). The statistical heterogeneity of the evidence was moderate (P<.001; I^2=58%). The quality of the evidence was very low as it was downgraded by 6 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5).

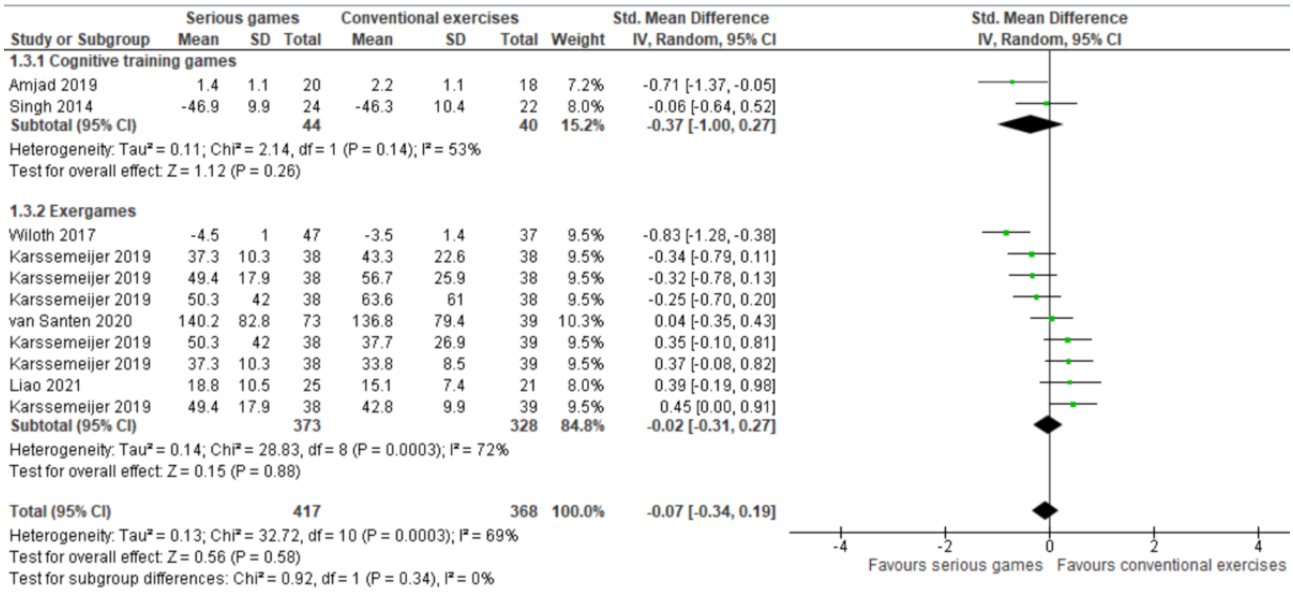
Two types of serious games were used in this comparison (ie, serious games vs conventional exercises): cognitive training games and exergames. We conducted a subgroup analysis to investigate whether cognitive training games and exergames had a different effect on processing speed (Figure 5 [38-43]). The subgroup analysis of 12% (2/16) of the studies showed no statistically significant difference (P=.26) in processing speed between the cognitive training game group and the conventional exercise group (SMD: −0.37, 95% CI −1.00 to 0.27). The statistical heterogeneity of the evidence was moderate (P=.14; I^2=53%). The quality of the evidence was very low as it was downgraded by 4 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5). Furthermore, the subgroup analysis of 25% (4/16) of the studies (9 comparisons) showed no statistically significant difference (P=.88) in processing speed between the exergame group and the conventional exercise group (SMD: −0.02, 95% CI −0.31 to 0.27). The statistical heterogeneity of the evidence was substantial (P<.001; I^2=72%). The quality of the evidence was very low as it was downgraded by 6 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 5).

Figure 4. Forest plot of 6 studies (11 comparisons) comparing the effect of serious games on processing speed with that of conventional exercises [38-43].



XSL•FO
RenderX

Figure 5. Forest plot of 6 studies (11 comparisons) comparing the effect of cognitive training games and exergames on processing speed with that of conventional exercises [38-43].



Serious Games Versus Other Serious Games

In total, 19% (3/16) of the studies assessed the effect of serious games on processing speed in comparison with that of other serious games [44-46]. The first study (1/3, 33%) compared the effect of a cognitive training game with that of exergames and found no statistically significant difference (*P*=.76) in processing speed between the groups [46]. The remaining 67% (2/3) of the studies compared the effect of cognitive training games that adjust the level of difficulty of the tasks based on an individual's mastery in each level (ie, adaptive games) with that of the same games but without adjustment of the level of difficulty of the tasks (ie, nonadaptive games) [44,45]. Of the 2 studies, 1 (50%) showed no statistically significant difference in processing speed between the groups as measured by the Delis-Kaplan Executive Function System Color-Word Interference Test 1 (*P*=.91) and Delis-Kaplan Executive Function System Color-Word Interference Test 2 (*P*=.51) [44]. The remaining study (1/2, 50%) did not report the findings related to the outcome measure Trail Making Test A [45]. The first author of that study was contacted to obtain these findings, but he has not replied.

Discussion

Principal Findings

This study compiled evidence regarding the effectiveness of serious games in improving processing speed among older adults with cognitive impairment. Our review revealed that serious games are as effective as no or passive interventions and conventional exercises in improving processing speed and that there is no difference between cognitive training games and exergames when compared with conventional exercises. The nonsuperior effect of serious games over no or passive interventions and conventional exercises may have been due to the following reasons: (1) serious games in the included studies targeted cognitive abilities in general rather than processing speed specifically and (2) the sample size in most of the included studies (10/16, 62%) was small (<100).

Our findings are consistent with findings of a previous review that showed no statistically significant difference in processing speed between the cognitive training game group and the passive or active intervention group among older adults with MCI or dementia [17]. However, our findings are not consistent with findings of previous reviews [16,18-20]. Specifically, 4 reviews showed that cognitive training games [16,19,20] and exergames [18] are more effective than other interventions (passive or active) in improving processing speed among healthy older adults. This inconsistency may be attributed to the fact that (1) all these reviews focused on older adults without cognitive impairment only, whereas our review focused on older adults with cognitive impairment [16,18-20]; (2) they included quasi-experiments and pilot RCTs [18,19]; and (3) they did not compare the effect of serious games with that of a specific comparator (eg, no intervention, conventional exercises, or conventional cognitive activities) [16,18-20].

Strengths and Limitations

Strengths

This study is the only one of the aforementioned reviews [16-20] that compared both serious games and their types with a specific comparator (ie, no intervention and conventional exercises) and used the GRADE approach to assess the quality of the evidence that resulted from the meta-analyses, thereby enabling the reader to draw more accurate conclusions. Given that we included only RCTs, which are the most rigorous research method for studying cause-effect relationships [47], our findings are more likely to be reliable than findings generated from reviews that included other study designs such as pilot RCTs and quasi-experiments.

The risk of publication bias in this review was minimal given that we sought to identify as many relevant studies as possible through (1) searching the most popular databases in the IT and health fields and gray literature databases, (2) conducting backward and forward reference list checking, and (3) using a well-developed search query. The risk of selection bias in this review was not a concern as the study selection, data extraction,

XSL•FO
RenderX

risk of bias assessment, and quality of evidence appraisal were conducted by 2 reviewers independently.

Limitations

This review focused on the effectiveness of digital serious games in improving processing speed among older adults with cognitive impairment. Hence, this review cannot comment on the effectiveness (1) of nondigital serious games or those used for nontherapeutic purposes (eg, diagnosis), (2) of serious games in improving other cognitive abilities (eg, attention, learning, and memory), and (3) of serious games among other age groups or those without cognitive impairment.

The effect size estimated for each meta-analyzed study was likely overestimated or underestimated as we used postintervention data rather than the pre-post intervention change data to calculate it. We used postintervention data as most studies (12/16, 75%) did not report the mean and SD for pre-post intervention change in processing speed for each group, and the difference in processing speed between groups at baseline was not statistically significant in all studies.

This review focused only on the short-term effect of serious games by pooling only postintervention data rather than follow-up data given that the follow-up period was not consistent among the 44% (7/16) of studies that reported follow-up data. As a result, we are unable to speculate on the long-term impact of serious games on processing speed. Given that this review did not include research published before 2010, studies written in a language other than English, quasi-experiments, or pilot RCTs, it is likely that this review missed some relevant studies.

Practical and Research Implications

Practical Implications

The findings of this review should be cautiously interpreted for the following reasons: (1) the quality of the evidence from all meta-analyses was very low mainly because of high risk of bias, high heterogeneity, and imprecision of the estimated total effect sizes; (2) the number of studies included in some meta-analyses was small; and (3) the sample sizes in many meta-analyzed studies were small. Consequently, until more robust evidence is available, serious games should be offered or used as a supplement rather than an alternative intervention targeting processing speed.

None of the included studies used smartphones as a platform for serious games. Smartphones are more appealing than other platforms as they are less expensive, more accessible, and more pervasive than computers and gaming consoles. In 2021, the global number of mobile devices and users was estimated to be approximately 15 billion and 7.1 billion, respectively, with these statistics likely to climb dramatically by 2025 [48]. Thus, we recommend that gaming companies develop serious games that can be played via smartphones. None of the serious games in the included studies were designed to target processing speed. Therefore, there is an urgent need to develop serious games that specifically target processing speed.

Research Implications

Although this review addressed the research gap related to the short-term effect of serious games on processing speed among older adults with cognitive impairment, the following research gaps need further reviews to be bridged: (1) the long-term effect of serious games, (2) the effect of serious games on other cognitive abilities (eg, attention and visuospatial skills) and other disorders (eg, attention-deficit/hyperactivity disorder, autism, and motor disabilities), and (3) the effect of serious games among people from different age groups with or without cognitive impairment.

As mentioned earlier, most studies in this review did not report the mean and SD for pre-post intervention change in processing speed. Researchers should report such information in their future publications to enable reviewers to calculate a more accurate effect size for each study. According to a previous review [18], there have been many studies conducted to assess the effect of exergames on processing speed among healthy older adults. However, only 25% (4/16) of the studies in this review investigated the effect of exergames on processing speed among older adults with cognitive impairment. Further trials are needed to address the aforementioned research gap.

In this review, only 6% (1/16) of the trials compared serious games with conventional cognitive training, and only 12% (2/16) compared adaptive serious games with nonadaptive serious games. We urge researchers to examine the aforementioned comparisons to reach more definitive conclusions. None of the included studies investigated the effect of serious games that specifically target processing speed rather than cognitive abilities in general. Future studies should use serious games that specifically target processing speed to examine their effect.

The overall risk of bias was low in only 19% (3/16) of the included trials as the remaining studies had issues mainly in the randomization process or selection of the reported results (ie, unpublished protocol or analysis plan). Future trials should improve their quality by minimizing such bias. To this end, they should be conducted and reported according to recommended guidelines or tools such as the Risk of Bias 2 [27], and they should have a large sample size that is enough to obtain the desired statistical power. As most of the included studies (14/16, 88%) were conducted in high-income countries, the generalizability of the findings of this review to lower-income countries may be restricted owing to the diversity of their cultures and socioeconomic conditions. Thus, more trials should be conducted in lower-income countries.

Conclusions

Serious games did not have a superior effect on processing speed among older adults with cognitive impairment in comparison with no or passive interventions and conventional exercises. However, this finding should be cautiously interpreted for the following reasons: (1) the quality of the evidence from all meta-analyses was very low mainly because of high risk of bias, high heterogeneity, and imprecision of the estimated total effect sizes; (2) the number of studies included in some meta-analyses was small; and (3) the sample sizes in many meta-analyzed studies were small. Therefore, until more robust evidence is

XSL•FO
RenderX

available, serious games should be offered or used as a supplement rather than as an alternative intervention targeting processing speed. Future reviews should investigate the long-term impact of serious games on other cognitive abilities and disorders among people from different age groups with or without cognitive impairment. Further trials should be undertaken to investigate the effect of serious games that specifically target processing speed rather than cognitive abilities in general.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 26 KB - games_v10i3e36754_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 30 KB - games_v10i3e36754_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 25 KB - games_v10i3e36754_app3.docx]

Multimedia Appendix 4
Reviewers' judgments about each "risk of bias" domain for each included study.
[DOCX File , 108 KB - games_v10i3e36754_app4.docx]

Multimedia Appendix 5
Grading of Recommendations Assessment, Development, and Evaluation profile for comparison of serious games with control and conventional exercises regarding processing speed.
[DOC File , 18 KB - games_v10i3e36754_app5.doc]

References

1. Ageing and health. World Health Organization. 2021 Oct 4. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2022-01-16]
2. Mild Cognitive Impairment (MCI). Alzheimer's Association. URL: https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment [accessed 2022-01-16]
3. Alzheimer's Association. 2015 Alzheimer's disease facts and figures. Alzheimers Dement 2015 Mar;11(3):332-384. [doi: 10.1016/j.jalz.2015.02.003] [Medline: 25984581]
4. Cognitive impairment: a call for action, now. Centers for Disease Control Prevention. 2011 Feb. URL: https://www.cdc.gov/aging/pdf/cognitive_impairment/cogimp_poilicy_final.pdf [accessed 2022-08-24]
5. Landhuis E. Detecting Alzheimer's gets easier with a simple blood test. Scientific American. 2021 Feb 4. URL: https://www.scientificamerican.com/article/detecting-alzheimers-gets-easier-with-a-simple-blood-test/# [accessed 2022-03-23]
6. Nilsson L. Memory function in normal aging. Acta Neurol Scand Suppl 2003;179:7-13. [doi: 10.1034/j.1600-0404.107.s179.5.x] [Medline: 12603244]
7. Salthouse TA. Memory aging from 18 to 80. Alzheimer Dis Assoc Disord 2003;17(3):162-167. [doi: 10.1097/00002093-200307000-00008] [Medline: 14512830]
8. Coppin A, Shumway-Cook A, Saczynski J, Patel K, Ble A, Ferrucci L, et al. Association of executive function and performance of dual-task physical tests among older adults: analyses from the InChianti study. Age Ageing 2006 Nov;35(6):619-624 [FREE Full text] [doi: 10.1093/ageing/afl107] [Medline: 17047008]
9. Royall D, Palmer R, Chiodo L, Polk M. Declining executive control in normal aging predicts change in functional status: the Freedom House Study. J Am Geriatr Soc 2004 Mar;52(3):346-352. [doi: 10.1111/j.1532-5415.2004.52104.x] [Medline: 14962147]
10. Yakhno NN, Zakharov VV, Lokshina AB. Impairment of memory and attention in the elderly. Neurosci Behav Physiol 2007 Mar;37(3):203-208. [doi: 10.1007/s11055-007-0002-y] [Medline: 17294094]
11. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev 1996 Jul;103(3):403-428. [doi: 10.1037/0033-295x.103.3.403] [Medline: 8759042]

XSL•FO
RenderX

12. Salthouse TA, Ferrer-Caja E. What needs to be explained to account for age-related effects on multiple cognitive variables? Psychol Aging 2003 Mar;18(1):91-110. [doi: 10.1037/0882-7974.18.1.91] [Medline: 12641315]
13. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
14. Jaramillo-Alcázar A, Luján-Mora S, Salvador-Ullauri L. Inclusive education: mobile serious games for people with cognitive disabilities. Enfoque UTE 2018 Mar 30;9(1):53-66. [doi: 10.29019/enfoqueute.v9n1.236]
15. Abd-Alrazaq A, Alajlani M, Alhuwail D, Toro C, Giannicchi A, Ahmed A, et al. The effectiveness and safety of serious games for improving cognitive abilities among elderly people with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Mar 10;10(1):e34592 [FREE Full text] [doi: 10.2196/34592] [Medline: 35266877]
16. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
17. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
18. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
19. Wang G, Zhao M, Yang F, Cheng LJ, Lau Y. Game-based brain training for improving cognitive function in community-dwelling older adults: a systematic review and meta-regression. Arch Gerontol Geriatr 2021;92:104260. [doi: 10.1016/j.archger.2020.104260] [Medline: 32980574]
20. Bonnechère B, Langley C, Sahakian BJ. The use of commercial computerised cognitive games in older adults: a meta-analysis. Sci Rep 2020 Sep 17;10(1):15276 [FREE Full text] [doi: 10.1038/s41598-020-72281-3] [Medline: 32943742]
21. Kletzel SL, Sood P, Negm A, Heyn PC, Krishnan S, Machtinger J, et al. Effectiveness of brain gaming in older adults with cognitive impairments: a systematic review and meta-analysis. J Am Med Dir Assoc 2021 Nov;22(11):2281-8.e5 [FREE Full text] [doi: 10.1016/j.jamda.2021.05.022] [Medline: 34146517]
22. Yen H, Chiu H. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
23. Abd-Alrazaq A, Alhuwail D, Al-Jafar E, Ahmed A, Shuweihdi F, Reagu SM, et al. The effectiveness of serious games in improving memory among older adults with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Aug 09;10(3):e35202 [FREE Full text] [doi: 10.2196/35202] [Medline: 35943792]
24. Abd-Alrazaq A, Alhuwail D, Ahmed A, Househ M. Effectiveness of serious games for improving executive functions among older adults with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Jul 25;10(3):e36123 [FREE Full text] [doi: 10.2196/36123] [Medline: 35877166]
25. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
26. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
27. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Deeks J, Higgins J, Altman D. Analysing data and undertaking meta‐analyses. In: Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Sussex, UK: John Wiley & Sons; Sep 22, 2008.
29. Schunemann H, Oxman A, Vist G, Higgins J, Deeks J, Glasziou P. Interpreting results and drawing conclusions. In: Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Sussex, UK: John Wiley & Sons; Sep 22, 2008.
30. Landis JR, Koch GG. An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977 Jun;33(2):363. [doi: 10.2307/2529786]
31. Finn M, McDonald S. Repetition-lag training to improve recollection memory in older people with amnestic mild cognitive impairment. A randomized controlled trial. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2015;22(2):244-258. [doi: 10.1080/13825585.2014.915918] [Medline: 24820545]
32. Robert P, Manera V, Derreumaux A, Ferrandez Y Montesino M, Leone E, Fabre R, et al. Efficacy of a web app for cognitive training (MeMo) regarding cognitive and behavioral performance in people with neurocognitive disorders: randomized controlled trial. J Med Internet Res 2020 Mar 11;22(3):e17167 [FREE Full text] [doi: 10.2196/17167] [Medline: 32159519]
33. Savulich G, Piercy T, Fox C, Suckling J, Rowe JB, O'Brien JT, et al. Cognitive training using a novel memory game on an iPad in patients with amnestic mild cognitive impairment (AMCI). Int J Neuropsychopharmacol 2017 Aug 01;20(8):624-633 [FREE Full text] [doi: 10.1093/ijnp/pyx040] [Medline: 28898959]

XSL•FO
RenderX

34. Valdes EG, O'Connor ML, Edwards JD. The effects of cognitive speed of processing training among older adults with psychometrically- defined mild cognitive impairment. Curr Alzheimer Res 2012 Nov;9(9):999-1009. [doi: 10.2174/156720512803568984] [Medline: 22594383]
35. Yang Y, Kwak YT. Improvement of cognitive function after computer-based cognitive training in early stage of Alzheimer's dementia. Dement Neurocogn Disord 2017 Mar;16(1):7-11 [FREE Full text] [doi: 10.12779/dnd.2017.16.1.7] [Medline: 30906364]
36. Thapa N, Park HJ, Yang J, Son H, Jang M, Lee J, et al. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. J Clin Med 2020 Apr 29;9(5):1283 [FREE Full text] [doi: 10.3390/jcm9051283] [Medline: 32365533]
37. Tarnanas I, Tsolakis A, Tsolaki M. Assessing virtual reality environments as cognitive stimulation method for patients with MCI. In: Technologies of Inclusive Well-Being. Berlin, Heidelberg: Springer; 2014.
38. Fiatarone Singh MA, Gates N, Saigal N, Wilson GC, Meiklejohn J, Brodaty H, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. J Am Med Dir Assoc 2014 Dec;15(12):873-880. [doi: 10.1016/j.jamda.2014.09.010] [Medline: 25444575]
39. Amjad I, Toor H, Niazi IK, Pervaiz S, Jochumsen M, Shafique M, et al. Xbox 360 kinect cognitive games improve slowness, complexity of EEG, and cognitive functions in subjects with mild cognitive impairment: a randomized control trial. Games Health J 2019 Apr;8(2):144-152. [doi: 10.1089/g4h.2018.0029] [Medline: 30239213]
40. Wiloth S, Werner C, Lemke NC, Bauer J, Hauer K. Motor-cognitive effects of a computerized game-based training method in people with dementia: a randomized controlled trial. Aging Ment Health 2018 Sep 06;22(9):1124-1135. [doi: 10.1080/13607863.2017.1348472] [Medline: 28682124]
41. van Santen J, Dröes RM, Twisk JW, Blanson Henkemans OA, van Straten A, Meiland FJ. Effects of exergaming on cognitive and social functioning of people with dementia: a randomized controlled trial. J Am Med Dir Assoc 2020 Dec;21(12):1958-67.e5 [FREE Full text] [doi: 10.1016/j.jamda.2020.04.018] [Medline: 32651132]
42. Karssemeijer EG, Aaronson JA, Bossers WJ, Donders R, Olde Rikkert MG, Kessels RP. The quest for synergy between physical exercise and cognitive stimulation via exergaming in people with dementia: a randomized controlled trial. Alzheimers Res Ther 2019 Jan 05;11(1):3 [FREE Full text] [doi: 10.1186/s13195-018-0454-z] [Medline: 30611286]
43. Liao Y, Chen I, Hsu W, Tseng H, Wang R. Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: a randomised controlled trial. Ann Phys Rehabil Med 2021 Sep;64(5):101492. [doi: 10.1016/j.rehab.2021.101492] [Medline: 33454398]
44. Flak MM, Hol HR, Hernes SS, Chang L, Engvig A, Bjuland KJ, et al. Adaptive computerized working memory training in patients with mild cognitive impairment. A randomized double-blind active controlled trial. Front Psychol 2019;10:807 [FREE Full text] [doi: 10.3389/fpsyg.2019.00807] [Medline: 31031677]
45. Hyer L, Scott C, Atkinson MM, Mullen CM, Lee A, Johnson A, et al. Cognitive training program to improve working memory in older adults with MCI. Clin Gerontol 2016;39(5):410-427. [doi: 10.1080/07317115.2015.1120257] [Medline: 29471774]
46. Park J, Park J. Does cognition-specific computer training have better clinical outcomes than non-specific computer training? A single-blind, randomized controlled trial. Clin Rehabil 2018 Feb;32(2):213-222. [doi: 10.1177/0269215517719951] [Medline: 28726492]
47. Bhattacherjee A. Social Science Research Principles, Methods, and Practices. Athens, GA, USA: Global Text Project; 2012.
48. Mobile Statistics Report, 2021-2025. The Radicati Group. 2021 Jan. URL: https://www.radicati.com/wp/wp-content/uploads/2021/Mobile_Statistics_Report,_2021-2025_Executive_Summary.pdf [accessed 2022-05-13]

Abbreviations

AD: Alzheimer disease
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
MCI: mild cognitive impairment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

XSL•FO
RenderX

Edited by N Zary; submitted 29.01.22; peer-reviewed by E Al-Jafar, M Alajlani, M Khalifa; comments to author 16.03.22; revised version received 23.03.22; accepted 07.04.22; published 09.09.22.

Please cite as:
Abd-alrazaq A, Ahmed A, Alali H, Aldardour AM, Househ M
The Effectiveness of Serious Games on Cognitive Processing Speed Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis
JMIR Serious Games 2022;10(3):e36754
URL: https://games.jmir.org/2022/3/e36754
doi:10.2196/36754
PMID:36083623

©Alaa Abd-alrazaq, Arfan Ahmed, Haitham Alali, Ahmad Mohammad Aldardour, Mowafa Househ. Originally published in JMIR Serious Games (https://games.jmir.org), 09.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prediction of Specific Anxiety Symptoms and Virtual Reality Sickness Using In Situ Autonomic Physiological Signals During Virtual Reality Treatment in Patients With Social Anxiety Disorder: Mixed Methods Study

Joo Young Chun[1*], MS; Hyun-Jin Kim[2*], MD; Ji-Won Hur[3], PhD; Dooyoung Jung[4], MD, PhD; Heon-Jeong Lee[5], MD, PhD; Seung Pil Pack[6], PhD; Sungkil Lee[7], PhD; Gerard Kim[8], PhD; Chung-Yean Cho[9], MFA; Seung-Moo Lee[9], MFA; Hyeri Lee[9], BA; Seungmoon Choi[10], PhD; Taesu Cheong[1], PhD; Chul-Hyun Cho[5], MD, PhD

[1]School of Industrial and Management Engineering, Korea University, Seoul, Republic of Korea
[2]Department of Psychiatry, Chungnam National University Sejong Hospital, Sejong, Republic of Korea
[3]School of Psychology, Korea University, Seoul, Republic of Korea
[4]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[5]Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea
[6]Department of Biotechnology and Bioinformatics, Korea University, Sejong, Republic of Korea
[7]Department of Software, Sungkyunkwan University, Suwon, Republic of Korea
[8]Department of Computer Science and Engineering, Digital Experience Laboratory, Korea University, Seoul, Republic of Korea
[9]Department of Film & Multimedia, Korea National University of Arts, Seoul, Republic of Korea
[10]Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Chul-Hyun Cho, MD, PhD
Department of Psychiatry
Korea University College of Medicine
73 Goryeodae-ro, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 029205505
Fax: 82 0262805810
Email: david0203@gmail.com

Abstract

Background: Social anxiety disorder (SAD) is the fear of social situations where a person anticipates being evaluated negatively. Changes in autonomic response patterns are related to the expression of anxiety symptoms. Virtual reality (VR) sickness can inhibit VR experiences.

Objective: This study aimed to predict the severity of specific anxiety symptoms and VR sickness in patients with SAD, using machine learning based on in situ autonomic physiological signals (heart rate and galvanic skin response) during VR treatment sessions.

Methods: This study included 32 participants with SAD taking part in 6 VR sessions. During each VR session, the heart rate and galvanic skin response of all participants were measured in real time. We assessed specific anxiety symptoms using the Internalized Shame Scale (ISS) and the Post-Event Rumination Scale (PERS), and VR sickness using the Simulator Sickness Questionnaire (SSQ) during 4 VR sessions (#1, #2, #4, and #6). Logistic regression, random forest, and naïve Bayes classification classified and predicted the severity groups in the ISS, PERS, and SSQ subdomains based on in situ autonomic physiological signal data.

Results: The severity of SAD was predicted with 3 machine learning models. According to the F1 score, the highest prediction performance among each domain for severity was determined. The F1 score of the ISS mistake anxiety subdomain was 0.8421

XSL•FO
RenderX

using the logistic regression model, that of the PERS positive subdomain was 0.7619 using the naïve Bayes classifier, and that of total VR sickness was 0.7059 using the random forest model.

Conclusions: This study could predict specific anxiety symptoms and VR sickness during VR intervention by autonomic physiological signals alone in real time. Machine learning models can predict the severe and nonsevere psychological states of individuals based on in situ physiological signal data during VR interventions for real-time interactive services. These models can support the diagnosis of specific anxiety symptoms and VR sickness with minimal participant bias.

Trial Registration: Clinical Research Information Service KCT0003854; https://cris.nih.go.kr/cris/search/detailSearch.do/13508

(JMIR Serious Games 2022;10(3):e38284) doi:10.2196/38284

KEYWORDS

social anxiety; virtual reality; autonomic physiological signals; machine learning; virtual reality sickness

Introduction

Anxiety involves uncertainty about the expectancy of a threat. It is a normal emotional response that helps individuals heed and cope with potential signs of danger [1]. Anxiety induces anticipatory stress and various neurophysiological responses. Previous research has shown that certain brain regions that process threatening information can biologically trigger stress response systems, such as the autonomic nervous system and hypothalamic-pituitary-adrenal axis. Hyperactivity of the autonomic response can induce the expression of anxiety signals, perception of sensations from inside the body, and central periphery interactions [2].

Social anxiety disorder (SAD) is the fear of negative evaluation and embarrassment in social situations. For instance, anxiety can occur when thinking about being observed while eating and speaking publicly [3]. The physiological symptoms of SAD include heart palpitations, sweating, tremors, shaking, and blushing. Many existing studies have focused on anxiety-induced heart reactions, and have discovered that anxiety increases blood catecholamine concentrations [4] and causes excess sympathetic activation and parasympathetic withdrawal [5].

Existing treatment methods for SAD include pharmacotherapy and psychotherapy, such as cognitive behavioral therapy (CBT) [6]. During CBT, an exposure technique allows the patient to experience and participate in a feared situation to create a natural process related to fear reduction [7]. However, individuals with social anxiety rarely request for support from professionals owing to fear of the stigma of mental health treatment. Therefore, the demand for accessibility to CBT via online content has been increasing. Moreover, the range of virtual reality (VR) treatments is expanding [8].

Recent VR intervention research has created virtual environments, such as restaurants. Individuals undergoing VR intervention showed reduced anxiety regarding daily social interactions. They also showed a decrease in social interaction anxiety and depression, and improved satisfaction with daily life after the VR intervention [9].

Exposure therapy is the mainstay VR treatment for SAD. Measuring anxiety symptoms in real time while applying the exposure technique enables the confirmation of actual SAD symptom levels and the appropriate application of personalized therapeutic methods.

However, VR sickness influences user experience when using VR systems [10]. The symptoms of VR sickness include eye fatigue, disorientation, and nausea [11]. Sickness resulting from VR is referred to by many names, including motion sickness, cybersickness, and simulator sickness. These uncomfortable feelings disturb VR intervention; therefore, VR sickness is a crucial problem that must be solved. The most popular method to assess VR sickness is the Simulator Sickness Questionnaire (SSQ) [12]. The symptoms of simulator motion sickness resemble those of motion sickness, but are less severe. Three distinct symptom groups, namely nausea, oculomotor, and disorientation, were clustered across 16 symptom variables. Because simulator sickness–related symptoms differ from motion sickness patterns, a measurement system was established through this grouping. The SSQ scores based on factor analysis models provide a scale score as a good indicator of overall simulator sickness severity. This quantifies simulator sickness for activities that can lead to symptoms. However, the method cannot be used for real-time measurements because it measures psychological scales through a questionnaire after VR intervention. There are several technical approaches for overcoming VR sickness, and measuring and evaluating it in real time is crucial. This is one of the factors determining the difference in individual sensitivity and future compliance. If it is not addressed during VR intervention, patient compliance can be greatly reduced [13].

Several studies have proposed interventions for social anxiety using VR treatment technology. The results showed that social interaction anxiety and quality of life improved when treatment was applied [9]. In particular, previous studies have shown that exposure to scenarios of various social situations through VR is effective for alleviating social anxiety symptoms [14].

Recently, a study applied VR treatment to anxiety disorders, such as SAD [15], proposed a VR intervention for SAD treatment, and verified its clinical effect by reducing scores on the Social Anxiety Disorder Scale. This method included CBT to enable participants to overcome SAD by exposing them to a self-introduction situation. The treatment had exposure therapy for patients with SAD through VR for over 6 sessions. Compared with previous studies that applied relatively unidirectional VR content, this study applied a customized approach that divided the difficulty according to the patient's

XSL•FO
RenderX

own symptom report. However, this study used a subjective self-assessment scale, which was influenced by patient bias. Furthermore, survey-oriented progress is lengthy and may have different results based on subjective judgment.

We have attempted to investigate SAD in relation to VR treatment using objective data such as brain imaging data [16]. Objective physiological signals, which reflect the degree of anxiety symptoms or sickness symptoms, can be assessed in real time during VR treatment sessions by several sensors. This would provide a basis for stand-alone interactive content to be provided during VR treatment for anxiety. Moreover, reducing VR motion sickness could help enhance user experience and compliance using physiological signals.

In this study, we developed a machine-learning model to predict the severity of specific anxiety symptoms and VR sickness using physiological signal data that were measured in situ during VR treatment sessions in patients with SAD.

Methods

Participants

Participants were recruited through advertisements on the universities' online sites. The inclusion criteria for the SAD group were set as follows: (1) men and women aged between 19 and 31 years using the Korean language; (2) condition meeting the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria for SAD, as evaluated by the Mini-International Neuropsychiatric Interview [17]; (3) no psychiatric comorbidity, except for depression and panic disorder, and no experience in psychotropic drug treatment; (4) not currently receiving psychotherapy; (5) not currently diagnosed with medical or neurological disorders; (6) no history of psychotic symptoms that can be triggered by VR interventions; and (7) no susceptibility to visual stimuli. We set the exclusion criteria as follows: (1) any history of organic brain damage or intellectual disability; (2) a history of psychotic symptoms that can be triggered by VR interventions; (3) vulnerability to visual stimuli; and (4) ineligibility for participation in a magnetic resonance imaging assessment (during the study, we performed a functional magnetic resonance imaging evaluation for another project in the same subjects) [18].

Forty individuals with SAD participated in the study. Among them, 8 individuals quit for personal reasons (eg, time limitations). Ultimately, 32 participants completed the study. All participants completed all VR sessions. The number of people who participated in the entire session was 32, but 6 of them, for whom the sensor data for a specific session were not collected because of problems such as sensor errors, were excluded from data analysis. Consequently, data from 26 individuals were used for specific anxiety symptom and VR sickness predictions after the data qualification process.

Ethics Approval

The study was approved by the Institutional Review Board of Korea University Hospital (2018AN0377). It was conducted in accordance with the principles of the Declaration of Helsinki. All participants were informed of the study procedure and provided written informed consent before the experiment. This study was registered on the Clinical Research Information Service (KCT0003854).

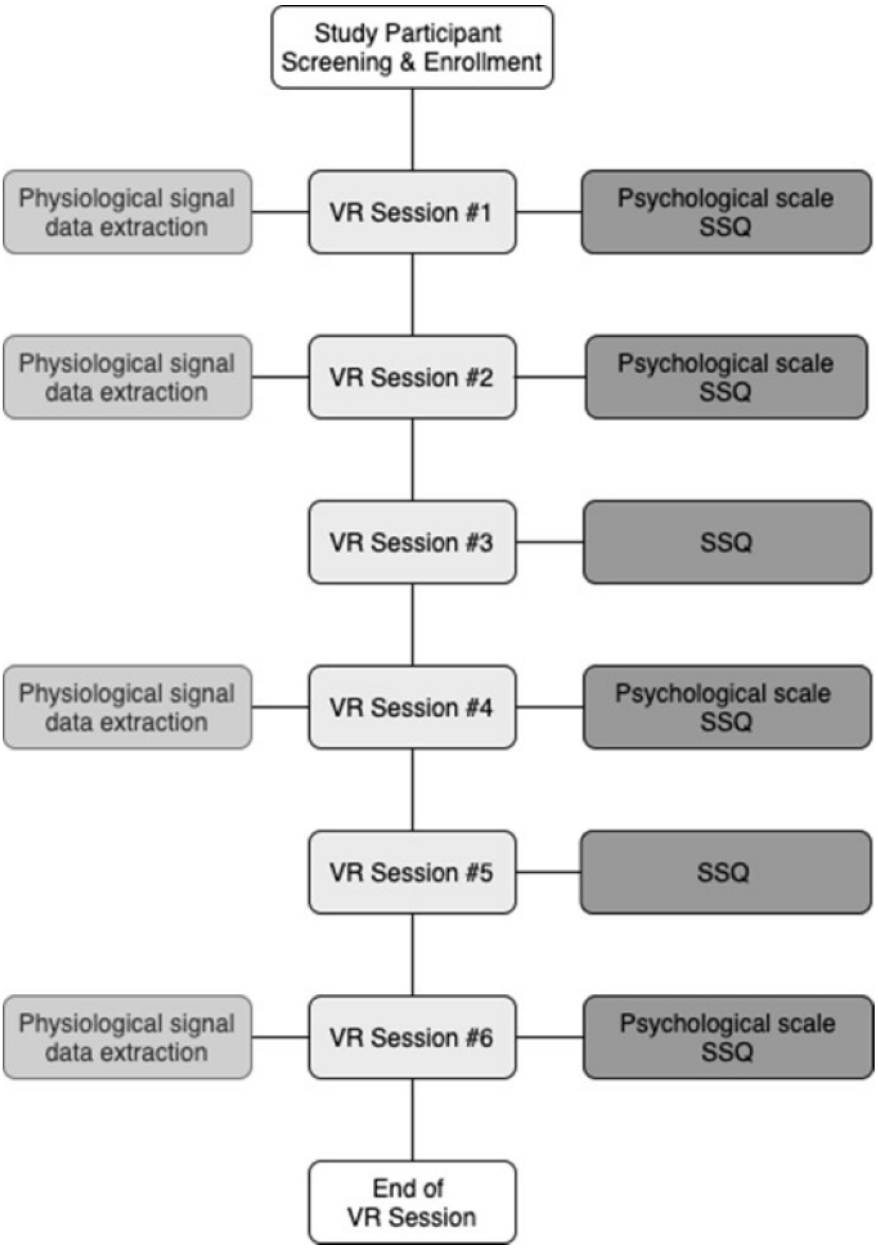
Study Design

VR Intervention for SAD

The study design included VR sessions, psychological scale evaluations, and in situ autonomic physiological signal data measurements (Figure 1). A total of 6 VR sessions and 4 psychological scale assessments were performed. A recent study found that the amount of anxiety that increased during actual conversation and the anxiety that increased during a conversation in a VR environment was similar [19]. In a study comparing VR exposure therapy and real-world exposure therapy for SAD [20], 97 participants with a major diagnosis of SAD were randomly assigned to VR exposure therapy, real-world exposure therapy, or a waiting list. A standardized self-report scale showed a statistically significant improvement in those who completed active treatment than in those on the waiting list. Thus, VR exposure therapy was as effective as real-world exposure therapy for treating SAD. In this study, participants completed VR treatment sessions that included interactive content on social anxiety situations, and the VR intervention was performed face-to-face and provided individually. The intervention proceeded in a VR environment, and each participant entered a meeting room with several nonplaying characters [15]. The content had introduction, core, and finishing stages. In the introduction phase, participants selected their avatars and were informed how to manipulate the VR device to progress through the stages [21]. Participants performed a warm-up session with mediation to support their adaptation. The core stage had an intervention in which college student's participated in social situations where they introduced themselves to each other. Participants proceeded with the treatment session by selecting the difficulty of the core level as easy, medium, or hard. In the first session, all participants started the VR intervention at the easy level. From the second session, participants freely chose the level of difficulty they wanted based on the difficulty they experienced during the treatment process. This is related to the degree of difficulty in interacting in social situations. The reaction of the nonplaying characters was based on the difficulty level selected in the previous stage. In the easy-level session, the nonplaying characters concentrated well on the patients' self-introduction. Nonplaying characters appeared distracted in the hard level, such as yawning or engaging in chatting. The finishing stage provided cognitive and behavioral safety guidelines for SAD with text and audio information during the VR intervention. The VR headset used was VIVE (HTC Corporation). The in situ autonomic physiological signals (heart rate [HR] and galvanic skin response [GSR]) of the participants were measured during the VR experience.

XSL•FO
RenderX

Figure 1. Virtual reality (VR) treatment process with psychological scale measurement and physiological signal data extraction. Psychological scale assessments included Beck Anxiety Inventory, State-Trait Anxiety Inventory, Social Phobia Scale, Social Interaction Anxiety Scale, Brief-Fear of Negative Evaluation Scale, Internalized Shame Scale, Post-Event Rumination Scale, and Liebowitz Social Anxiety Scale. SSQ: Simulator Sickness Questionnaire.



Data Collection and Separation

This study analyzed specific anxiety symptoms and in situ autonomic physiological signal data. For this purpose, physiological signal data were measured in real time during 4 VR sessions (#1, #2, #4, and #6), where psychological scale tests were performed after VR treatment. Data were extracted as independent variables. The SSQ was evaluated during all treatment sessions; therefore, in situ physiological signal data for all sessions were extracted. We created a data frame with the HR and GSR data. Values below zero in the HR and GSR measurements were considered outliers. Outliers were removed, and a moving average was used to reduce the noise of the data.

Measurement

Psychological Scales

We measured psychological scales using subjective surveys after the VR sessions #1, #2, #4, and #6. The Beck Anxiety Inventory [22], State-Trait Anxiety Inventory [23], Social Phobia Scale [24], Social Interaction Anxiety Scale [25], Brief-Fear of Negative Evaluation Scale [26], Internalized Shame Scale (ISS) [27], and Post-Event Rumination Scale (PERS) [28] were used. After building a simple linear regression model for the scales, their subdomains, and the SSQ, the best performing group of scales based on the mean squared error of the models' prediction results was analyzed. Subsequently, we generated a machine learning model to predict the survey results. The linear regression coefficient was included as an independent variable for prediction; therefore, the specific anxiety symptoms were

XSL•FO
RenderX

selected in the order of the smallest root mean square error deviation when predicting with simple linear regression. The subdomains of the corresponding specific anxiety symptoms were also included and used for prediction. The ISS and PERS showed a good reflection of the physical response to SAD. In this study, the machine learning model used the ISS and PERS as independent variables. Multimedia Appendix 1 provides the complete regression results for our machine learning model.

The ISS assesses the patients' shame and self-esteem [27]. It has 30 items that measure shame and self-esteem (24 and 6 items, respectively). ISS data were measured using the ISS written in the Korean language [29]. Factor analysis involved the following 4 factors: inadequacy, emptiness, self-punishment, and fear of mistakes, with 10, 5, 5, and 4 items, respectively.

The PERS measures postevent ruminations during social situations [28,30]. It comprises 2 scales, negative and positive rumination, with 5 and 9 items, respectively. Each answer was measured on a numerical scale from 0 (low score) to 4 (high score), with a higher score indicating that the individual frequently experiences rumination.

VR Sickness Scale

The SSQ was developed for VR sickness, and is the gold standard for assessing physical sickness after exposure to a simulator or VR environment [12]. The SSQ can measure 16 symptoms of VR sickness. The subdomains are grouped into nausea (general discomfort, increased salivation, sweating, nausea, difficulty concentrating, stomach awareness, and burping), oculomotor (general discomfort, fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, and blurred vision), and disorientation (difficulty focusing, nausea, fullness of the head, blurred vision, dizziness with eyes open, dizziness with eyes closed, and vertigo). The total VR sickness score was the sum of nausea, oculomotor, and disorientation group scores, with a higher score indicating a greater level of simulator sickness.

Data Labeling and Processing

We used k=2 to categorize the specific anxiety symptoms in the severe and nonsevere groups, where k is the number of median points of the clusters to be classified. Data are grouped as per the number of hyperparameters k. In k-means clustering, the algorithm can be executed when the user determines the number of clusters. K-means clustering is a representative detached clustering algorithm that uses unsupervised learning [31].

We labeled the severe and nonsevere groups using k-means clustering for each specific anxiety symptom and VR sickness. The ISS cutoff values for dividing the severe groups were 48 (total score), 11 (mistake anxiety), 9 (self-punishment), 10 (emptiness), and 16 (inappropriate). The PERS cutoff values were 41 (total score), 17 (positive), and 27 (negative). For the SSQ, severity classification was performed using k-means clustering via the sum of the response scores for each category. The cutoff values for dividing the severe group were 9 (total score), 7 (nausea), 8 (oculomotor), and 4 (disorientation).

Physiological Signal Data

We measured all participants' HR and GSR during the VR sessions in real time. These values are closely related to in situ physiological signaling responses related to anxiety. We recorded the data using a Shimmer3 GSR+ Unit (Shimmer) with 3 channels. Two channels tracked the electrodermal skin conductance level signals via hand electrodes attached to the first and middle fingers. We monitored the fingers of the nondominant hand at a sampling frequency of 52 Hz. The third channel recorded cardiac volume data from an infrared sensor attached to the left earlobe. Data were converted into HR data using a software system.

Preprocessing removed any sections where the HR and GSR were not measured or had negative values. The noise was reduced using the moving average method at 2-s intervals. We extracted all the data from the time series data set of the participants. We also calculated the mean, standard deviation, minimum, maximum, linear regression coefficient, maximum difference, peak ratio, and mean difference for each in situ physiological signal. These created the data frame. To calculate the peak ratio [32], we divided the number of peaks by the length of the physiological signal data.

We used real data for the prediction model; therefore, generating data was a challenge. The severe and nonsevere groups varied in size; therefore, we manipulated the data set to improve the model performance. We did not consider the undersampling technique because the number of participants was insufficient. Conversely, we used the oversampling technique on the minority group to improve classifier performance.

The SMOTE (Synthetic Minority Oversampling Technique) approach generated insufficient data [33]. This method helps construct classifiers from imbalanced data sets (where the classification categories are not approximately equally distributed), such as real-world data sets with a small number of abnormal cases.

Statistical Analysis

Group-Based Analysis

Differences between groups (severe and nonsevere) were assessed using independent *t* tests in Python with SciPy version 1.5.2. Differences were considered statistically significant at $P<.05$. Furthermore, we analyzed correlations between the extracted variables. The variables with the strongest positive correlations were mean and peak GSR ratios. The variables with the second strongest positive correlations were the mean and standard deviation of the GSR. The variables with the strongest negative correlations were the linear regression coefficient and standard deviation of the GSR. The second most negatively correlated variables were the GSR difference and HR linear regression coefficient. All correlations are illustrated in Multimedia Appendix 2. There was a high correlation between the mean and peak GSR (0.92); however, removing this feature decreased the prediction score. Therefore, this feature was retained. All other correlations were <0.86.

XSL•FO
RenderX

Machine Learning Techniques

All machine learning techniques were performed using Python version 3.7.7. Logistic regression [34], random forest [35], and naïve Bayes [36] classification were performed using Scikit-learn version 0.23.2. The random forest model avoids overfitting, thereby improving model accuracy. Additionally, we can select relatively important ranks of variables in the classification model. The naïve Bayes model is a conditional probability-based classification method that calculates the probability of features belonging to each class and handles noise and missing values reasonably.

In addition, when training, it works excellently regardless of the size of the data, and it is easy to obtain the estimated probabilities for prediction. Although 32 samples of data could be insufficient, we amplified the patients' data using SMOTE methodology after segmentation for each session. We used machine learning models that can perform reasonable prediction with a few data points to minimize performance degradation because of insufficient data. We calculated the accuracy, F1 score (weighted average of precision and recall), and area under the curve (AUC) for the prediction.

All machine learning models, except for the random forest model, were performed using the oversampling technique on the training data set. This supplemented insufficient data after the training-test split. We performed 5-fold cross-validation for the random forest model to improve the prediction score. During the prediction, sampling of the severe and nonsevere groups was stratified for training-test splits to ensure equally frequent severity.

Results

Classified Specific Anxiety Symptoms and VR Sickness

Specific anxiety symptom and VR sickness results were compared between the severe and nonsevere groups using k-means clustering. The smallest difference between group sizes was observed for ISS self-punishment (severe vs nonsevere group, 53 vs 51). Based on the average score of the specific anxiety symptoms, the most significant difference between the groups was the ISS score (severe vs nonsevere group, 59.951 vs 33.222). ISS emptiness showed the smallest difference between groups (severe vs nonsevere group, 12.91 vs 5.75). After labeling with k-means clustering, the anxiety symptom with the lowest distortion was ISS mistake anxiety and that with the highest distortion was total ISS (1.263 vs 7.439).

After 4 VR sessions with 26 participants, 104 sets of VR sickness data were collected. These were labeled as severe and nonsevere groups using k-means clustering. Distortion of k-means clustering was 2.57 for total VR sickness, 1.44 for nausea, 1.71 for oculomotor, and 1.16 for disorientation. Table 1 presents the statistical analysis results of the severe and nonsevere groups. The boxplot for the severe and nonsevere groups is illustrated in Multimedia Appendix 3.

Table 1. Severe and nonsevere groups clustered by k-means clustering.

Measure	Groups[a]				Minimum value	Cutoff value	Maximum value	Distortion[b]
	Nonsevere group count	Severe group count	Nonsevere group mean	Severe group mean				
ISS[c]	63	41	33.222	59.951	3	<48	92	7.439
ISS mistake anxiety	34	70	8.176	12.957	1	<11	16	1.263
ISS self-punishment	51	53	4.961	12.453	1	<9	20	1.835
ISS emptiness	60	44	5.750	12.910	0	<10	20	2.104
ISS inappropriate	58	46	9.155	21.978	2	<16	36	3.344
PERS[d]	47	57	32.766	48.140	19	<41	64	4.203
PERS positive	36	68	10.059	21.000	2	<17	35	3.007
PERS negative	50	54	16.900	36.016	3	<27	56	5.163
Total VR[e] sickness	73	31	3.247	13.710	0	<9	29	2.568
Nausea group	98	6	1.786	10.500	0	<7	15	1.445
Oculomotor group	63	41	2.238	8.341	0	<6	15	1.709
Disorientation group	69	35	0.986	6.000	0	<4	12	1.164

[a]After labeling into severe and nonsevere groups through k-means clustering, the numerical characteristics and differences between the groups are shown for each group.

[b]Distortion was calculated using the k-means clustering model with k=2.

[c]ISS: Internalized Shame Scale.

[d]PERS: Post-Event Rumination Scale.

[e]VR: virtual reality.

XSL•FO
RenderX

Differences in Physiological Signals According to Specific Anxiety Symptom Severity

There were significant differences in the GSR min between the severe and nonsevere groups stratified by the ISS (t_{102}=1.39; *P*=.17). In addition, groups stratified by the ISS mistake anxiety showed a significant difference in the HR peak ratio (t_{102}=2.07; *P*=.04) and GSR average change (t_{102}=2.02; *P*=.046). Groups stratified by ISS self-punishment showed significantly different GSR min values (t_{102}=2.34; *P*=.02). Groups stratified by ISS emptiness showed a difference in mean HR (t_{102}=−2.44; *P*=.02). Groups stratified by ISS inappropriate showed a significant difference in mean HR (t_{102}=−2.23; *P*=.03). Groups stratified by PERS had significantly different mean HRs (t_{102}=−1.99; *P*=.050). Furthermore, groups stratified by PERS positivity showed significantly different mean HRs (t_{102}=2.51; *P*=.01), HR min values (t_{102}=2.49; *P*=.01), and HR peak ratios (t_{102}=2.15; *P*=.03). Groups stratified by PERS negative had significantly different mean HRs (t_{102}=−2.79; *P*=.006), HR min values (t_{102}=−2.49; *P*=.01), HR peak ratios (t_{102}=−2.15; *P*=.03), and GSR max values (t_{102}=−2.00; *P*=.048). The severe and nonsevere groups by anxiety symptoms are illustrated in Multimedia Appendix 4.

We performed *t* tests after labeling the severity of VR sickness, for each VR sickness subdomain. The most significant difference in total VR sickness was for the HR min value (t_{102}=−1.63; *P*=.02). In the nausea group, the HR max value was significantly different between the groups (t_{102}=−2.47; *P*=.02). Additionally, the oculomotor group differed in the HR min value (t_{102}=−2.19; *P*=.03). The disorientation group showed a difference in total HR change (t_{102}=−2.14; *P*=.04) and total GSR change (t_{102}=−2.09; *P*=.04). Differences in the severe and nonsevere groups according to VR sickness are illustrated in Multimedia Appendix 5.

Specific Anxiety Symptom and VR Sickness Prediction Based on the Physiological Signal Data

After oversampling the minority data using SMOTE, the F1 score [37], accuracy, and area under the receiver operating characteristic curve were calculated after predicting the survey results using logistic regression analysis, random forest, and naïve Bayesian methodology. Additionally, we calculated the feature importance of specific anxiety symptoms. The variable importance can be calculated based on the impurity index [38]. We used the feature importance of a random forest–based prediction model.

Anxiety Symptom Prediction

The classifications of specific anxiety symptom results were analyzed based on 3 models. Logistic regression [34], random forest [35], and naïve Bayes classifier [36] analyzed the severity of specific anxiety symptoms for participants. The data were amplified using the SMOTE method to prevent performance degradation due to unbalanced data.

The classification performance was measured using the F1 score. The F1 scores of the ISS mistake anxiety subdomain were 0.8421, 0.7368, and 0.7647, as calculated by logistic regression, random forest, and naïve Bayesian classifier, respectively. These values were higher than those of the other ISS subdomains. The classification performance of the PERS negative subdomain was relatively higher than that of the PERS positive subdomain, with 0.6667, 0.6154, and 0.6452 as measured by logistic regression, random forest, and naïve Bayes classifier, respectively.

Specifically, we classified the ISS using a logistic regression model with an F1 score of 0.7619. The logistic regression model revealed an F1 score of 0.8421 for ISS mistake anxiety, whereas the naïve Bayes classification model revealed a score of 0.7857 for ISS self-punishment. The logistic regression model revealed an F1 score of 0.6429 for ISS emptiness and 0.7200 for ISS inappropriate. The random forest model revealed an F1 score of 0.7568 for PERS. Further, the naïve Bayes classification model showed an F1 score of 0.7619 for PERS positivity. Finally, the random forest model revealed an F1 score of 0.7097 for PERS negativity. The specific anxiety symptom prediction results are illustrated in Table 2 and Multimedia Appendix 6.

For ISS, ISS mistake anxiety, ISS self-punishment, ISS inappropriate, PERS, PERS positive, and PERS negative, physiological signal data related to GSR had the highest importance, and for ISS emptiness, physiological signal data related to HR had the highest importance. The feature importance of specific anxiety symptoms in the random forest model is illustrated in Figure 2.

XSL•FO
RenderX

Table 2. Specific anxiety symptom classification model evaluation with F1 score, accuracy, and area under the curve.

Variable[a]	Logistic regression			Random forest			Naïve Bayes classifier		
	F1 score	Accuracy	AUC[b]	F1 score	Accuracy	AUC	F1 score	Accuracy	AUC
ISS[c]	0.762	0.808	0.827	0.632	0.731	0.706	0.727	0.769	0.775
ISS mistake anxiety	0.842	0.769	0.733	0.790	0.692	0.608	0.765	0.692	0.660
ISS self-punishment	0.750	0.692	0.673	0.727	0.769	0.769	0.786	0.769	0.769
ISS emptiness	0.643	0.615	0.625	0.571	0.654	0.639	0.600	0.692	0.673
ISS inappropriate	0.720	0.731	0.732	0.571	0.654	0.639	0.714	0.692	0.702
PERS[d]	0.667	0.654	0.676	0.757	0.654	0.625	0.667	0.615	0.607
PERS positive	0.615	0.615	0.630	0.500	0.769	0.660	0.762	0.808	0.861
PERS negative	0.667	0.654	0.654	0.710	0.654	0.654	0.645	0.577	0.577

[a]After predicting the severity of each specific anxiety symptom using logistic regression, random forest, and naïve Bayes classifier models, the performance of each model was evaluated.

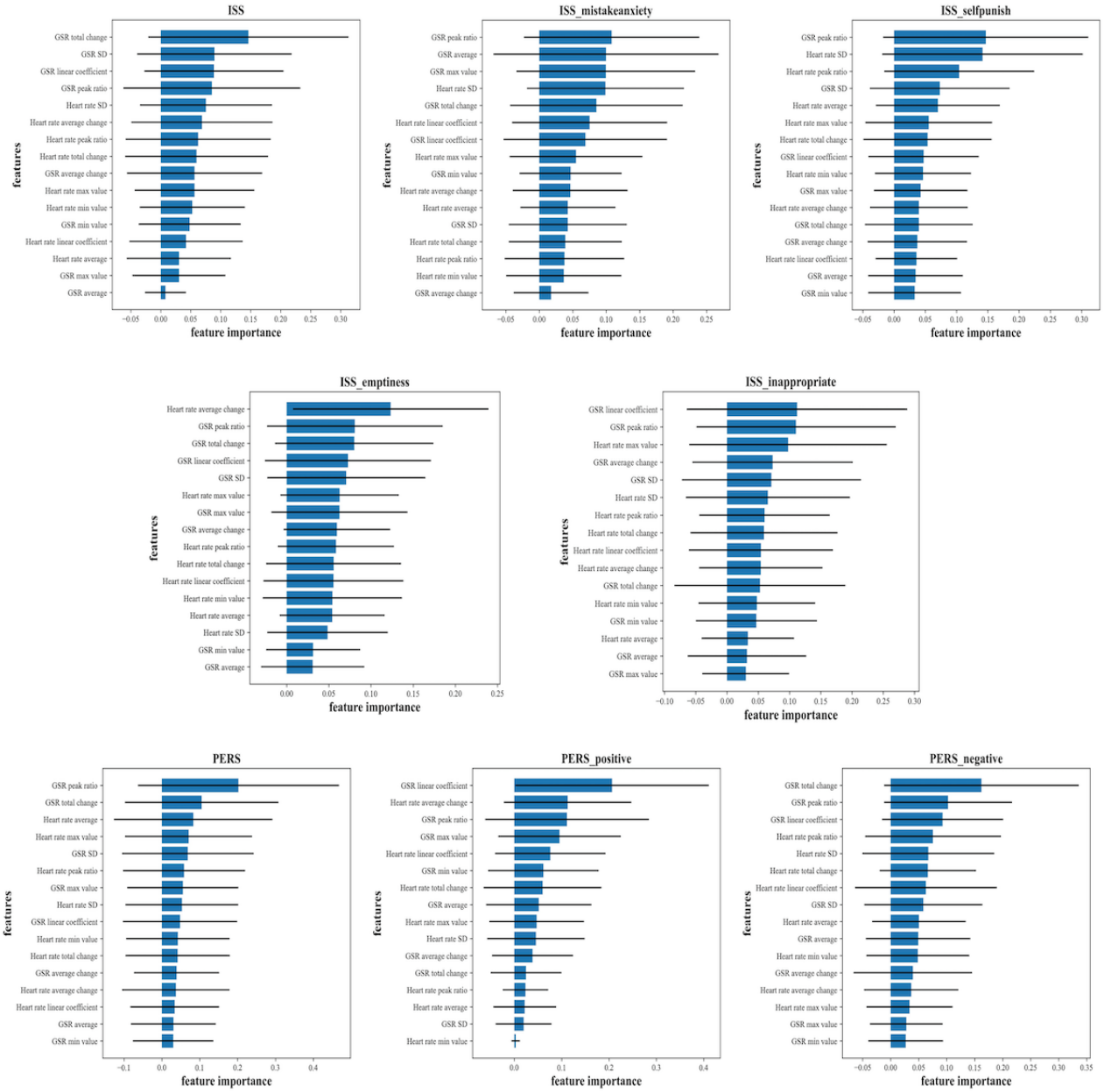
[b]AUC: area under the curve.

[c]ISS: Internalized Shame Scale.

[d]PERS: Post-Event Rumination Scale.

XSL•FO
RenderX

Figure 2. Feature importance of specific anxiety symptoms (random forest model). After predicting the subdomains of each anxiety symptom using the random forest model, the feature importance for each model was calculated and sorted in descending order. GSR: galvanic skin response; ISS: Internalized Shame Scale; PERS: Post-Event Rumination Scale.



VR Sickness Prediction

We classified the severity of total VR sickness with an F1 score of 0.7059, that in the nausea group with a score of 0.4000, that in the oculomotor group with a score of 0.6667, and that in the disorientation group with a score of 0.6364. The classification performance for the nausea group subdomain had the highest AUC of 0.94. The VR sickness results are illustrated in Table 3 and Multimedia Appendix 7.

When feature importance was calculated for the classification of the severity of VR sickness for each type, physiological signal data related to GSR were selected as essential variables for all subdomains. All subdomains of VR sickness symptoms showed HR-related features as the second most crucial factor in variation. Figure 3 illustrates the importance of VR sickness features from the random forest model.

XSL•FO
RenderX

Table 3. Virtual reality sickness classification model evaluation with F1 score, accuracy, and area under the curve.

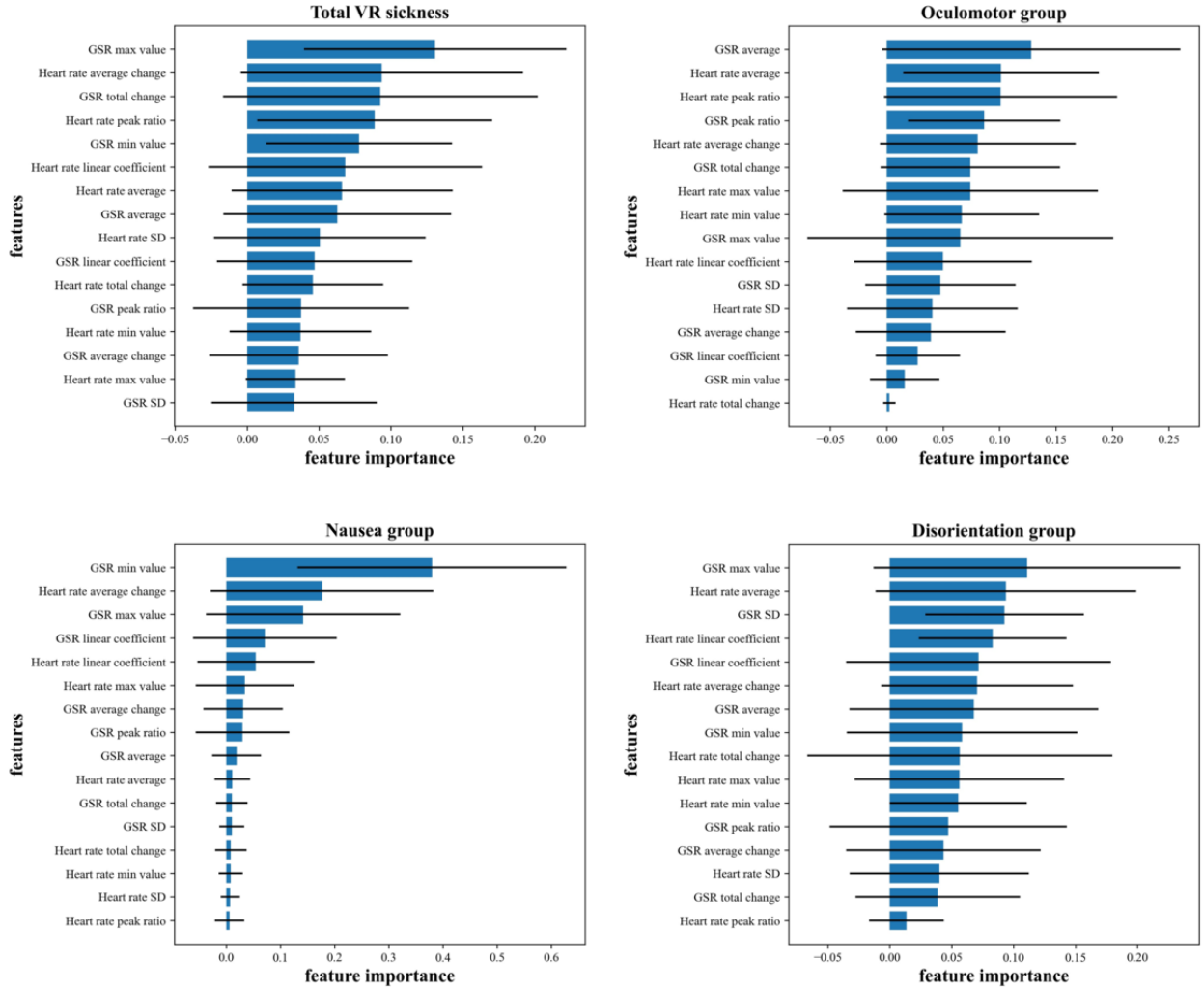
Variable[a]	Random forest		
	F1 score	Accuracy	AUC[b]
Total VR[c] sickness	0.7059	0.8077	0.7917
Nausea group	0.4000	0.8846	0.9400
Oculomotor group	0.6667	0.6538	0.7000
Disorientation group	0.6364	0.6923	0.7124

[a]After predicting the severity of each VR sickness scale using the random forest model, the model's performance was evaluated.

[b]AUC: area under the curve.

[c]VR: virtual reality.

Figure 3. Feature importance of virtual reality (VR) sickness (random forest model). After predicting the subdomains of each VR sickness scale using the random forest model, the feature importance for each model was calculated and sorted in descending order. GSR: galvanic skin response.



Discussion

Principal Findings

We developed and tested machine learning models to predict specific anxiety symptoms and VR sickness for SAD using in situ autonomic physiological signal data measured during participatory and interactive VR treatment. The severity of specific anxiety symptoms and the side effects of VR treatment are an essential part of digital therapy. Usually, these kinds of evaluations rely on subjective reporting by patients. However, autonomic physiological responses, such as blushing, sweating, and shivering, could play a central role in assessing symptoms [39]; therefore, they have the potential to be used for monitoring various symptoms related to anxiety or VR sickness.

Previous studies have assessed the relationship between physiological cues and SAD. However, real-time analysis is limited because of current sensor technology and evaluation

XSL•FO
RenderX

methods [40]. Furthermore, individuals with SAD cannot be treated because of methodological problems [41-43].

VR treatment techniques extend beyond the traditional psychiatric therapeutic approach [44]. To create a therapeutic VR system, we must build an interactive system rather than simply provide a VR environment. Interactive VR may have several benefits for psychiatric disorders in terms of treatment. First, it enhances ecological validity by immersing individuals in emotionally engaging virtually constructed therapeutic environments. Second, it can be flexibly used to present patients in various contexts, enabling personalized treatment according to the assessment of individual differences in symptoms during VR treatment [45]. We expect our findings to be useful in interactive VR treatment, especially for specific anxiety symptoms and VR sickness.

We showed that the data extracted from the time series with different in situ autonomic physiological signal data lengths could be used as the independent variables of a predictive model. As 32 samples of physiological data were not sufficient for the machine learning model, we increased the data by dividing it per session using only 32 participants' intervention data and supplementing the insufficient data using SMOTE methodology. This technique can minimize performance degradation due to a lack of data. We generated cutoff values for the classification groups for labeling via k-means clustering. This is an unsupervised learning method based on the absence of a label to evaluate the model performance. We could explain which variable plays the most crucial role in predicting specific anxiety symptoms and VR sickness.

Among the anxiety symptoms related to the ISS, the most predictable was ISS mistake anxiety. Emotion comprised 5 main subsystems, namely, "cognitive component (appraisal), neurophysiological component (bodily symptoms), motivational component (action tendencies), motor expression component (facial and vocal expression), and subjective feeling component (emotional experience)" [46]. Shame is highly correlated with the body's physiological responses because of its mechanism [47]. Mistake anxiety is the core symptom of SAD. The high predictability of these symptoms is related to the reliability of HR and GSR predictions of SAD symptoms.

Among the scales related to PERS, the most predictable was PERS positivity. Rumination may negatively affect physical and mental health [48,49]. However, this is inconsistent with previous studies reporting that positive rumination either is not associated with social anxiety or is low [28,30]. In contrast, other studies have reported that positive and negative rumination is high in social anxiety. Socially anxious individuals ruminate broadly, positively, or negatively about all aspects of social interactions when faced with ambiguity [50]. However, no previous study has examined the relationship between in situ autonomic physiological responses, and negative and positive rumination. Additional studies are required to determine whether these results are reproducible.

Several factors can cause VR sickness. Humans perceive their direction and movement through various sensory organs. Individuals may experience motion sickness if they repeatedly receive sensory information that differs from their prediction [51]. Visual movement may cause motion sickness [52,53]. When visual stimulation is the primary cause of motion sickness, it is called visually induced motion sickness. Previous studies have assessed virtual perceptions related to recognizing self-motion [54]. If the degree of physiological arousal (HR, blood pressure, skin conductance, respiration, skin temperature, and blood volume pulse amplitude) is high, the possibility of VR motion sickness is high [55].

We used HR and GSR for machine learning–based prediction in our study. These are important indicators of physiological arousal. SAD-related psychological symptoms and VR sickness during VR treatment can be determined in advance through the proposed model. The intervention of medical staff can also improve patient compliance. However, we must develop additional real-time measurement factors to better predict other VR sickness subdomain symptoms.

We predicted specific anxiety symptom severity and VR sickness severity via in situ physiological signal data from actual cases. We used supervised and unsupervised learning and data generation to build and evaluate the SAD and VR sickness predictive models. Considering these advantages, specific anxiety symptoms and VR sickness could be assessed more accurately. The data were labeled using an unsupervised learning method. After separating the severe and nonsevere groups, significant differences were found for each classification group using a *t* test. The real-time evaluation of VR motion sickness can help improve patient compliance with treatment. Moreover, we can reduce the time required for a survey by predicting participants' anxiety using objective data. Predicting survey results using objective in situ autonomic physiological signal data makes less subjective intervention possible compared with the conventional survey-based method. In addition, the evaluation of symptoms through real-time autonomic physiological signals shows the possibility of evaluating psychiatric symptoms in real time to increase interactivity, an essential element in the personalized VR treatment process.

Limitations

This study had several limitations. First, our study evaluated in situ autonomic physiological signals and symptom-related data in a SAD group alone. Subsequently, it included no control groups. This study used machine learning models to discriminate between severe and nonsevere groups because real-time physiological signal data were measured only for participants above a certain Korean Social Interaction Anxiety Scale (K-SAD) score. In some cases, after labeling the severe and nonsevere groups, there was no significant difference between groups. These results can be inferred as a limitation of the methodology that divides the group with a K-SAD score above 82 into severe and nonsevere groups according to specific anxiety symptoms. Although an unsupervised learning methodology distinguished respondents, it was challenging to explain clear criteria other than to the participants in the experiment. Second, this study used self-rating score-based scales, which could be associated with bias. Third, this study could only classify the severity of SAD (K-SAD ≥82) in the group with a particular score or higher. Fourth, there was insufficient evaluation of the response to a specific task or

XSL•FO
RenderX

exposure situation during a VR session. Because insufficient data are obtained using a SMOTE model, there is a limitation in not using the actual data. In addition, it is unknown how performance changes as an individual progresses during a session. Furthermore, we cannot determine how in situ autonomic physiological signals change when a specific situation occurs in each session. Fifth, the number of participants who participated in the VR treatment session was 32, but the number of participants used for the analysis was 26 (a small sample size). The decrease in performance can be inferred to be limited to the improvement in performance because the number of individuals who participated in the VR sessions was small. Therefore, we need to carefully interpret the results and secure a larger number of samples for future studies. However, since it is challenging to obtain real-time in situ data, we believe that the strength of this study is that real valuable data from participants were utilized for analysis.

Conclusion

This study showed that using in situ autonomous physiological signal data measured during a VR intervention can predict specific anxiety symptoms and VR sickness in patients with SAD. Using real-time physiological data from VR sessions, we can classify the severity of specific SAD symptoms and utilize the findings for personalized digital treatment. Machine learning models can assist in the decisions of medical staff and the construction of interactive VR treatment. Future research should focus on various predictive methodologies to enhance the tailored interactive function and to maximize the convenience of VR treatment. Additionally, to improve the clinical prediction performance and increase accuracy, more abundant and appropriate data will need to be collected.

Acknowledgments

This work was supported by a National Research Foundation (NRF) of Korea grant funded by the Ministry of Science and Information and Communications Technology (MSIT), the Government of Korea (NRF-2020R1C1C1007463 and NRF-2021R1A5A8032895), and Information and Communications Technology (ICT) and Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2022M3C1B6080866).

Authors' Contributions

Authors CHC (david0203@gmail.com) and TC (tcheong@korea.ac.kr) are co-corresponding authors for this article, responsible for data and materials, manuscript submission, peer review, publication process, authorship details, and ethics committee approval. JWH, DJ, HJL, SL, GK, CYC, SML, SC, TC, and CHC conceived and designed the study. JYC, HJK, TC, and CHC performed statistical analyses. JYC, HJK, JWH, DJ, TC, and CHC wrote the first draft of this manuscript. JYC, HJK, JWH, DJ, SL, GK, CYC, SML, HL, SC, TC, and CHC participated in data collection. All authors have edited all versions of the manuscript. All authors were involved in interpreting the results, and they read, commented on, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Psychological scale prediction (linear regression model).
[PDF File (Adobe PDF File), 344 KB - games_v10i3e38284_app1.pdf]

Multimedia Appendix 2
Correlation table for each feature.
[PDF File (Adobe PDF File), 298 KB - games_v10i3e38284_app2.pdf]

Multimedia Appendix 3
Box plot of specific anxiety symptoms and virtual reality sickness severe and nonsevere group clustering results.
[PDF File (Adobe PDF File), 864 KB - games_v10i3e38284_app3.pdf]

Multimedia Appendix 4
Results of specific anxiety symptoms.
[PDF File (Adobe PDF File), 385 KB - games_v10i3e38284_app4.pdf]

Multimedia Appendix 5
Results of virtual reality sickness.
[PDF File (Adobe PDF File), 380 KB - games_v10i3e38284_app5.pdf]

XSL•FO
RenderX

Multimedia Appendix 6
Receiver operating characteristic curve of specific anxiety symptom prediction models (logistic regression model, random forest model, and naïve Bayesian model).
[PDF File (Adobe PDF File), 853 KB - games_v10i3e38284_app6.pdf]

Multimedia Appendix 7
Receiver operating characteristic curve of a virtual reality sickness prediction model (random forest model).
[PDF File (Adobe PDF File), 382 KB - games_v10i3e38284_app7.pdf]

References

1. Cannistraro P, Rauch S. Neural circuitry of anxiety: evidence from structural and functional neuroimaging studies. Psychopharmacol Bull 2003;37(4):8-25. [Medline: 15131515]
2. Garfinkel SN, Eccles JA, Critchley HD. The heart, the brain, and the regulation of emotion. JAMA Psychiatry 2015 Nov;72(11):1071-1072. [doi: 10.1001/jamapsychiatry.2015.1493] [Medline: 26376236]
3. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
4. Thurston RC, Rewak M, Kubzansky LD. An anxious heart: anxiety and the onset of cardiovascular diseases. Prog Cardiovasc Dis 2013 May;55(6):524-537. [doi: 10.1016/j.pcad.2013.03.007] [Medline: 23621962]
5. Friedman BH. An autonomic flexibility-neurovisceral integration model of anxiety and cardiac vagal tone. Biol Psychol 2007 Mar;74(2):185-199. [doi: 10.1016/j.biopsycho.2005.08.009] [Medline: 17069959]
6. Rapee RM, Heimberg RG. A cognitive-behavioral model of anxiety in social phobia. Behaviour Research and Therapy 1997 Aug;35(8):741-756. [doi: 10.1016/s0005-7967(97)00022-3]
7. Heimberg RG. Cognitive-behavioral therapy for social anxiety disorder: current status and future directions. Biological Psychiatry 2002 Jan;51(1):101-108. [doi: 10.1016/s0006-3223(01)01183-0]
8. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113. [doi: 10.1097/hrp.0000000000000138]
9. Geraets CN, Veling W, Witlox M, Staring AB, Matthijssen SJ, Cath D. Virtual reality-based cognitive behavioural therapy for patients with generalized social anxiety disorder: a pilot study. Behav. Cogn. Psychother 2019 Mar 27;47(6):745-750. [doi: 10.1017/s1352465819000225]
10. McCauley ME, Sharkey TJ. Cybersickness: perception of self-motion in virtual environments. Presence: Teleoperators & Virtual Environments 1992 Jan;1(3):311-318. [doi: 10.1162/pres.1992.1.3.311]
11. LaViola JJ. A discussion of cybersickness in virtual environments. SIGCHI Bull 2000 Jan 01;32(1):47-56. [doi: 10.1145/333329.333344]
12. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
13. Sevinc V, Berkman MI. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. Appl Ergon 2020 Jan;82:102958. [doi: 10.1016/j.apergo.2019.102958] [Medline: 31563798]
14. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger É, Forget H, et al. Virtual reality compared with exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. Br J Psychiatry 2017 Apr;210(4):276-283. [doi: 10.1192/bjp.bp.116.184234] [Medline: 27979818]
15. Kim H, Lee S, Jung D, Hur J, Lee H, Lee S, et al. Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: longitudinal questionnaire study. J Med Internet Res 2020 Oct 06;22(10):e23024 [FREE Full text] [doi: 10.2196/23024] [Medline: 33021481]
16. Lee H, Choi J, Jung D, Hur J, Cho C. The effects of virtual reality treatment on prefrontal cortex activity in patients with social anxiety disorder: participatory and interactive virtual reality treatment study. J Med Internet Res 2021 Dec 17;23(12):e31844 [FREE Full text] [doi: 10.2196/31844] [Medline: 34801979]
17. Yoo S, Kim Y, Noh J, Oh K, Kim C, NamKoong K, et al. Validity of Korean version of the mini-international neuropsychiatric interview. Anxiety and Mood 2006;2(1):50-55 [FREE Full text]
18. Hur J, Shin H, Jung D, Lee H, Lee S, Kim GJ, et al. Virtual reality-based psychotherapy in social anxiety disorder: fMRI study using a self-referential task. JMIR Ment Health 2021 Apr 14;8(4):e25731 [FREE Full text] [doi: 10.2196/25731] [Medline: 33851931]
19. Powers MB, Briceno NF, Gresham R, Jouriles EN, Emmelkamp PMG, Smits JAJ. Do conversations with virtual avatars increase feelings of social anxiety? J Anxiety Disord 2013 May;27(4):398-403. [doi: 10.1016/j.janxdis.2013.03.003] [Medline: 23746490]

XSL•FO
RenderX

20. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, et al. Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. J Consult Clin Psychol 2013 Oct;81(5):751-760. [doi: 10.1037/a0033559] [Medline: 23796315]
21. Social Anxiety Disorder VR treatment. YouTube. URL: https://www.youtube.com/channel/UCCpzKK_qEqz8LRfxl_TfGXw/featured [accessed 2022-08-29]
22. Beck A, Epstein N, Brown G, Steer R. Beck anxiety inventory. J of Consult Clin Psychol 1993:23-40 [FREE Full text] [doi: 10.1037/t02025-000]
23. Spielberger CD, Gonzalez-Reigosa F, Martinez-Urrutia A, Natalicio LF, Natalicio DS. The state-trait anxiety inventory. Revista Interamericana De Psicología/Interamerican Journal of Psychology 1971;5:145-158 [FREE Full text]
24. Davidson J, Potts N, Richichi E, Ford S, Krishnan K, Smith R, et al. The Brief Social Phobia Scale. J Clin Psychiatry 1991 Nov;52 Suppl:48-51. [doi: 10.1037/t07672-000] [Medline: 1757457]
25. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]
26. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consult Clin Psychol 1969 Aug;33(4):448-457. [doi: 10.1037/h0027806] [Medline: 5810590]
27. Cook DR. Measuring shame. Alcoholism Treatment Quarterly 1988 Mar 07;4(2):197-215. [doi: 10.1300/j020v04n02_12]
28. Abbott MJ, Rapee RM. Post-event rumination and negative self-appraisal in social phobia before and after treatment. J Abnorm Psychol 2004 Mar;113(1):136-144. [doi: 10.1037/0021-843X.113.1.136] [Medline: 14992666]
29. Lee I, Choi H. Assessment of shame and its relationship with maternal attachment, hypersensitive narcissism and loneliness. Korean Journal of Counseling and Psychotherapy 2005;17(3):651-670 [FREE Full text]
30. Edwards S, Rapee R, Franklin J. Postevent rumination and recall bias for a social performance event in high and low socially anxious individuals. Cognitive Therapy and Research 2003;27:603-617. [doi: 10.1023/A:1026395526858]
31. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]
32. Duarte M. detecta: A Python module to detect events in data (Version v0.0.5). Zenodo. URL: https://zenodo.org/record/4598962#.YwzRTLRBzDc [accessed 2022-08-29]
33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
34. LaValley MP. Logistic regression. Circulation 2008 May 06;117(18):2395-2399. [doi: 10.1161/circulationaha.106.682658]
35. Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing 2007 Feb 22;26(1):217-222. [doi: 10.1080/01431160412331269698]
36. Murphy KP. Naive Bayes classifiers. University of British Columbia. 2006. URL: https://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes.pdf [accessed 2022-08-29]
37. Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. In: Sattar A, Kang B, editors. AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science, vol 4304. Berlin, Heidelberg: Springer; 2006:1015-1021.
38. Breiman L. Random forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
39. Gerlach AL, Mourlane D, Rist F. Public and private heart rate feedback in social phobia: a manipulation of anxiety visibility. Cogn Behav Ther 2004 Mar;33(1):36-45. [doi: 10.1080/16506070310014682] [Medline: 15224627]
40. Alvares GA, Quintana DS, Kemp AH, Van Zwieten A, Balleine BW, Hickie IB, et al. Reduced heart rate variability in social anxiety disorder: associations with gender and symptom severity. PLoS One 2013;8(7):e70468 [FREE Full text] [doi: 10.1371/journal.pone.0070468] [Medline: 23936207]
41. Rösler L, Göhring S, Strunz M, Gamer M. Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. J Behav Ther Exp Psychiatry 2021 Mar;70:101600 [FREE Full text] [doi: 10.1016/j.jbtep.2020.101600] [Medline: 32882674]
42. Wen W, Liu G, Mao Z, Huang W, Zhang X, Hu H, et al. Toward constructing a real-time social anxiety evaluation system: exploring effective heart rate features. IEEE Trans. Affective Comput 2020 Jan 1;11(1):100-110. [doi: 10.1109/taffc.2018.2792000]
43. Zhang X, Wen W, Liu G, Hu H. Recognition of public speaking anxiety on the recurrence quantification analysis of GSR signals. 2016 Presented at: Sixth International Conference on Information Science and Technology (ICIST); May 06-08, 2016; Dalian, China. [doi: 10.1109/icist.2016.7483472]
44. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019 Jul 19;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
45. Verhoef REJ, van Dijk A, Verhulp EE, de Castro BO. Interactive virtual reality assessment of aggressive social information processing in boys with behaviour problems: A pilot study. Clin Psychol Psychother 2021 May 17;28(3):489-499 [FREE Full text] [doi: 10.1002/cpp.2620] [Medline: 34048619]
46. Scherer KR. What are emotions? And how can they be measured? Social Science Information 2016 Jun 29;44(4):695-729. [doi: 10.1177/0539018405058216]

XSL•FO
RenderX

47. Hedman E, Ström P, Stünkel A, Mörtberg E. Shame and guilt in social anxiety disorder: effects of cognitive behavior therapy and association with social anxiety and depressive symptoms. PLoS One 2013;8(4):e61713 [FREE Full text] [doi: 10.1371/journal.pone.0061713] [Medline: 23620782]
48. Brosschot JF, Verkuil B, Thayer JF. Conscious and unconscious perseverative cognition: is a large part of prolonged physiological activity due to unconscious stress? J Psychosom Res 2010 Oct;69(4):407-416. [doi: 10.1016/j.jpsychores.2010.02.002] [Medline: 20846542]
49. Glynn LM, Christenfeld N, Gerin W. The role of rumination in recovery from reactivity: cardiovascular consequences of emotional states. Psychosom Med 2002;64(5):714-726. [doi: 10.1097/01.psy.0000031574.42041.23] [Medline: 12271102]
50. Dannahy L, Stopa L. Post-event processing in social anxiety. Behav Res Ther 2007 Jun;45(6):1207-1219. [doi: 10.1016/j.brat.2006.08.017] [Medline: 17112463]
51. Sherman C. Motion sickness: review of causes and preventive strategies. J Travel Med 2002;9(5):251-256 [FREE Full text] [doi: 10.2310/7060.2002.24145] [Medline: 12962599]
52. Bonato F, Bubka A, Palmisano S, Phillip D, Moreno G. Vection change exacerbates simulator sickness in virtual environments. Presence: Teleoperators and Virtual Environments 2008 Jun 01;17(3):283-292. [doi: 10.1162/pres.17.3.283]
53. Keshavarz B, Riecke BE, Hettinger LJ, Campos JL. Vection and visually induced motion sickness: how are they related? Front Psychol 2015 Apr 20;6:472 [FREE Full text] [doi: 10.3389/fpsyg.2015.00472] [Medline: 25941509]
54. Lubeck AJ, Bos JE, Stins JF. Motion in images is essential to cause motion sickness symptoms, but not to increase postural sway. Displays 2015 Jul;38:55-61. [doi: 10.1016/j.displa.2015.03.001]
55. Macedonio MF, Parsons TD, DiGiuseppe RA, Weiderhold BA, Rizzo AA. Immersiveness and physiological arousal within panoramic video-based virtual reality. Cyberpsychol Behav 2007 Aug;10(4):508-515. [doi: 10.1089/cpb.2007.9997] [Medline: 17711358]

Abbreviations

AUC: area under the curve
CBT: cognitive behavioral therapy
GSR: galvanic skin response
HR: heart rate
ISS: Internalized Shame Scale
K-SAD: Korean Social Interaction Anxiety Scale
PERS: Post-Event Rumination Scale
SAD: social anxiety disorder
SMOTE: Synthetic Minority Oversampling Technique
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by N Zary; submitted 27.03.22; peer-reviewed by H Kim, K Lee, K Uludag; comments to author 18.04.22; revised version received 30.06.22; accepted 21.07.22; published 16.09.22.

Please cite as:
Chun JY, Kim HJ, Hur JW, Jung D, Lee HJ, Pack SP, Lee S, Kim G, Cho CY, Lee SM, Lee H, Choi S, Cheong T, Cho CH
Prediction of Specific Anxiety Symptoms and Virtual Reality Sickness Using In Situ Autonomic Physiological Signals During Virtual Reality Treatment in Patients With Social Anxiety Disorder: Mixed Methods Study
JMIR Serious Games 2022;10(3):e38284
URL: https://games.jmir.org/2022/3/e38284
doi:10.2196/38284
PMID:36112407

©Joo Young Chun, Hyun-Jin Kim, Ji-Won Hur, Dooyoung Jung, Heon-Jeong Lee, Seung Pil Pack, Sungkil Lee, Gerard Kim, Chung-Yean Cho, Seung-Moo Lee, Hyeri Lee, Seungmoon Choi, Taesu Cheong, Chul-Hyun Cho. Originally published in JMIR Serious Games (https://games.jmir.org), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effects of a Computer Game (Healthy Rat King) on Preschool Children's Nutritional Knowledge and Junk Food Intake Behavior: Nonrandomized Controlled Trial

Ing-Chau Chang[1*], PhD; Cheng-Ying Yang[2*], PhD; Chin-En Yen[3*], PhD

[1]Department of Computer Science and Information Engineering, National Changhua University of Education, Changhua, Taiwan

[2]Department of Computer Science, University of Taipei, Taipei, Taiwan

[3]Department of Early Childhood Development and Education, Chaoyang University of Technology, Taichung, Taiwan

[*]all authors contributed equally

Corresponding Author:
Chin-En Yen, PhD
Department of Early Childhood Development and Education
Chaoyang University of Technology
168, Jifeng East Road
Wufeng District
Taichung, 41349
Taiwan
Phone: 886 4 23323000 ext 4706
Fax: 886 4 23742357
Email: ceyen@cyut.edu.tw

Abstract

Background: Playing computer-aided games could enhance children's interest in learning about nutritional knowledge and eventually promote healthy dietary intake behavior.

Objective: This study aims to evaluate the effectiveness of a computer game (Healthy Rat King) in improving the knowledge on nutrition and junk food intake among preschool children in Taiwan.

Methods: This was a quasi-experimental study that utilized the computer game Healthy Rat King as the nutrition education tool. We recruited 104 preschool children (aged 5-6 years) from preschools in central Taiwan, who were assigned to either the experimental group (n=56) or the control group (n=48). In the experimental group, a 1-hour computer-based educational game intervention was included in the course for 4 consecutive weeks. The control group did not receive this intervention.

Results: The level of nutritional knowledge for children in the experimental group was significantly higher than those in the control group after 4 weeks (*P*=.002). Furthermore, the frequency of consumption of chocolate, candies, and ice cream (high-calorie junk food) was reduced in the experimental group. There was also no significant difference in the consumption of candy and chocolate (*P*=.54), ice cream and ice pops (*P*=.21), cake (*P*=.92), biscuit (*P*=.98), soft drinks (*P*=.52), and fruit juice and sugary drinks (*P*=.31) between the 2 groups in the posttest.

Conclusions: Teaching using a computer game could improve children's nutritional knowledge. However, the intake frequency of junk food among children in the experimental group showed no significant difference from those in the control group.

(JMIR Serious Games 2022;10(3):e33137) doi:10.2196/33137

KEYWORDS

computer games; nutrition knowledge; junk food

Introduction

Recently, nutrition-related problems in children, such as obesity, unbalanced diet, and high intake of junk food, have been an area of extensive focus among parents, child nutrition experts, and teachers on early childhood education. It has been reported that 93% of children intake packaged food, 68% consume sugary drinks more than once a week, and 53% intake foods that are high in fat, salt, calories, or sugar at least once a day [1]. Besides, a previous survey [2] showed that 21% of children in

XSL•FO
RenderX

Taiwan drink sugary beverage at least once a day, while 89% of children eat junk food.

The preschool age (ie, age 3-6 years) is an important period to set or correct the eating habits of children. The formation of eating habits not only impacts the individual's food choices but also incurs the health, learning, and behavior problems [3]. Studies have shown that young children with malnutrition might have lower cognitive function scores, poor psychological development, reduced fine motor skills, and limited activity [3,4]. Importantly, the eating behavior (eg, food choices or motives) will impact the occurrence of diseases in the future. "Junk food" is unhealthy, as it is high in calories, fat, sugar, and salt. Further, it will increase the risk of malnutrition, obesity, cardiovascular diseases, high blood pressure, diabetes, and other chronic diseases [5]. Its regular consumption might also result in lower vitamins, minerals, and essential fatty acids that eventually affect the growth and development of children [3]. Therefore, it is very important to establish correct eating habits in the early childhood period.

For young children, nutrition education is an important tool in the preschool period. Implementation of nutrition education in preschools could well improve nutritional knowledge and eating habits of young children. Teachers in preschools conduct nutrition education sessions according to the interests and development of young children, and increase the participation and experience of young children in the teaching process [6].

Computer represents an important tool for learning. Computer games are novel and attractive to children. They gradually become their favorite because of their simple, convenient, vivid, interesting, entertaining, and challenging features [7]. Therefore, applying computer games as a teaching tool could help in the development of children's cognitive abilities, interactive relationships, and operational skills [8]. As computer games combine entertainment with knowledge, the concept of learning while playing might achieve the purpose of education. Besides, a computer game could help with behavior development, character building, and skill development for children [8].

Many studies have used computer games to understand the relationship between children's nutritional knowledge and eating behaviors [9-11]. Results of these studies have shown that computer games have a positive impact on children's dietary intake, nutritional knowledge, attitudes, behaviors, and activity level [9-11]. Several studies indicate that after playing video games for 1-6 weeks, the children have a positive attitude toward healthy eating, make healthy food choices, and reduce their sugar consumption; additionally, their attitudes on nutrition and physical activity improved. These results support the use of educational computer games as viable tools to help young children improve their food knowledge and dietary behaviors [10,11].

However, some results show that computer games could not change children's long-term eating behaviors. Although a short-term computer game intervention might improve children's nutritional knowledge, it does not sufficiently impact their nutritional knowledge and actual eating behavior in the long term [9].

Thus, computer games could have positive effects on nutritional knowledge, dietary intake, attitudes, and behaviors of children [9,12]. Computer game is also an appropriate nutrition education tool to improve children's attitudes toward food. One limitation of previous studies (eg, [9,11]) was that they primarily included children aged 8-13. As a result, less studies are available for preschool children. Nowadays, computer games are rarely used as a teaching tool in the preschools. We thus performed this study to evaluate whether computer games could enhance children's learning motivation, computer games included in teaching activities are effective, and computer game–related product development are appropriate for preschool teachers. We will also evaluate the effects of educational games on children's nutritional knowledge and junk food intake behaviors.

Methods

Participants

This was a quasi-experimental study. We recruited 104 preschool children (aged 5-6 years) from preschools in central Taiwan, who were assigned to either the experimental group (n=56) or the control group (n=48). In the experimental group, a 1-hour computer-based educational game intervention was included in the course for 4 consecutive weeks. The control group did not receive this intervention. Informed consent was obtained from the parents of the participating children.

Computer Game Design

For the Construct 2 course, an HTML5-based computer game (Healthy Rat King) was implemented as the auxiliary nutrition education tool for children in the experimental group. After installing the necessary software, the teacher just clicked the file named *index.html*, executed the computer game program, and then used the game to teach Construct 2.

Background of the Healthy Rat King Game

The Healthy Rat King game involves a situation in which a mouse hides from cats in the forest, while various foods appeared along the escape route to replenish its energy. If the mouse ate healthy foods, it earned 10 points and ran faster. By contrast, if it ate junk foods, it lost 5 points, gained "unhealthy" weight, and ran slower. The game gets over if the mouse is caught by the cat (usually with a low score) or if it reaches home without being caught (usually with a high score).

The goal of Healthy Rat King is to make young children understand foods that are healthy, while it also educates them about foods with high fat, high sugar, and high calorie content (junk food). Children can learn an important concept from this game: eating too much junk food can easily make one both fat and unhealthy. Thus, people should avoid eating too much junk food, and instead consume healthier food options.

Before performing the formal study, 12 children were selected for a pretest to evaluate the suitability of the computer game.

Learning Sheet

The nutritional knowledge status of all children was evaluated using the learning sheet before and 4 weeks after the intervention. This learning sheet was designed by the researcher

XSL•FO
RenderX

(C-YY) and verified by 3 experts according to the purpose of the research, the appropriateness of the content of the inspection topic, and the clarity of the questionnaire. The learning sheet includes 3 major questions. The first and second questions were used to examine the children's cognition on unhealthy and healthy food choices (ie, what are junk foods and healthy foods). The third question was used to examine children's cognition on the influence of junk food intake on the body (ie, the effects of eating junk food on the body, such as easy to getting sick or obese). The full score on the learning sheet is 100 points.

Questionnaire

The questionnaire contains the background information (including age, gender, parental education level, parental occupation), dietary surveys, status of junk food intake frequency among preschool children, and the influence of computer game interventions on preschool children's junk food intake (parent perceptions of the computer game teaching). The questionnaire was filled out by the parents of the preschool children. The reliability of the questionnaire is .90 (Cronbach α). The content validity of the questionnaire was reviewed by 3 experts.

Study Process

The experimental group received the computer game education for 1 hour once a week for 4 consecutive weeks. The content of the computer game education comprised 3 parts: children's learning motivation (eg, picture book performance, music rhythm, finger ballad), development activities (computer game learning), and integration activities (group discussion).

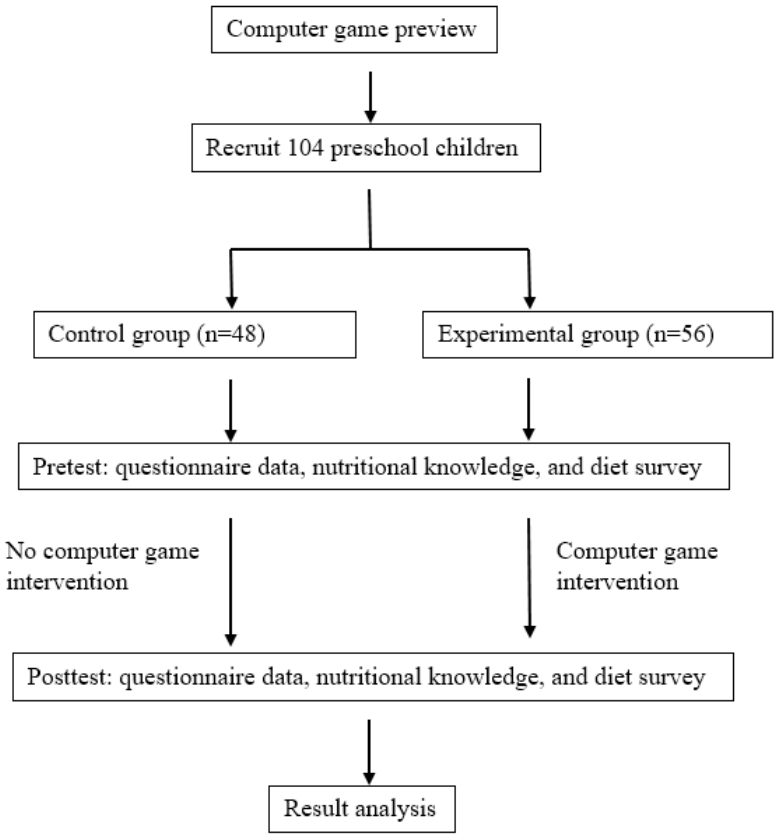
In the first stage, a pretest of the computer game is conducted to check teaching contents in 2 preschool classes.

In the second stage, all participating children have to complete the pretest of the learning sheet before the computer game course intervention.

In the third stage, the researchers conduct computer game education courses for the experimental group once an hour every week for 4 consecutive weeks. Each course content included the derived motivation, development activity (playing the computer game), and group discussion. Children in the experimental group were divided into subgroups (4-6 children per subgroup). They take turns to play computer games in the classroom. The playing game status is recorded by researchers. By contrast, the children in the control group did not receive the computer game course intervention. They performed general theme curriculum activities (derived motivation, art or music, or physical development activity) and participated in group discussions.

In the fourth stage, all participating children have to complete the posttest of the learning sheet after the computer game course intervention. The diet surveys and the junk food intake frequency questionnaire were also filled out by the parents of participating children. The study's flow diagram is illustrated in Figure 1.

Figure 1. Study flowchart.



XSL•FO
RenderX

Statistical Analysis

This study used SPSS for Windows (version 20.0; IBM, Inc.) for data analysis and statistics. Background data were subjected to descriptive statistics. The normality distributions of the data were analyzed by the Kolmogorov-Smirnov test, while a paired sample *t* test was used to compare the difference in nutritional knowledge scores and computer game scores between the pre- and posttest in the experimental group. It was also applied to compare the difference in nutritional knowledge scores between the pre- and posttest in the control group. The frequency of junk food intake among preschool children was described with the Bowker test of symmetry to compare the difference between the pre- and posttest. Differences in the frequency of junk food intake among preschool children between the experimental and control groups were determined using the chi-square test. The independent sample *t* test was used to compare the difference in nutritional knowledge scores between children in the experimental and control groups. A linear regression analysis was used to assess the relationship between the intervention (computer game course) and nutritional knowledge score after controlling for confounding factors (age, gender, parents education level, and parents occupation). A significance level of $P<.05$ was used for all tests.

Ethical Considerations

This study was exempt from research ethics review since it taught skills, and the effectiveness evaluation was conducted in a general teaching environment [13].

Results

Table 1 presents the demographic characteristics of the participants. Overall, there were more children with second birth ranking. Most fathers and mothers had a college degree.

Table 2 shows the computer game scores earned by the preschool children. The mean score was 25 (SD 23), 30 (SD 29), 38 (SD 33), and 37 (SD 25) in the first, second, third, and fourth weeks, respectively. Compared with the first week, a significant increase on game scores was noted in the third ($P=.02$) and fourth ($P<.001$) weeks. About 75% (42/56) of preschool children liked the computer game course *very much*, 13% (7/56) *liked* computer games, and only 5% (3/56) *disliked* the games.

Table 3 presents the nutritional knowledge scores of the preschool children. We used a learning sheet to evaluate the nutritional knowledge of preschool children. The mean pre- and posttest scores of the experimental group were significantly higher than those in the control group ($P=.26$ and $<.001$ for pre- and posttest, respectively). Furthermore, nutritional knowledge scores of the experimental group were significantly higher than those of the control group in the posttest ($P=.002$).

Table 4 shows the frequency of junk food intake between the experimental and control groups. Junk foods such as candies, soft drinks, cookies, cakes, packed snacks are unhealthy, as they are high in calories, sugar, and fat, but have little dietary fiber, protein, vitamin, minerals, or other important nutritional elements. There was no significant difference in the frequency of consumption of candy and chocolate ($P=.50$), ice cream and ice pops ($P=.48$), cake ($P=.60$), biscuit ($P=.72$), and soft drinks intake ($P=.70$) between the experimental and control groups; however, consumption of fruit juice and sugary drinks showed a significant difference between the experimental and control groups in the pretest ($P=.02$). There was also no significant difference in the consumption of candy and chocolate ($P=.54$), ice cream and ice pops ($P=.21$), cake ($P=.92$), biscuit ($P=.98$), soft drinks ($P=.52$), and fruit juice and sugary drinks ($P=.31$) between the 2 groups in the posttest.

Table 5 presents the relationship between the computer game course intervention and the nutritional knowledge score. The computer games intervention was significantly related to the improvement of children's nutritional knowledge ($P=.001$).

Following the intervention, the percentage of children who refused to eat junk food was significantly higher than those in the pretest ($P=.01$); additionally, the percentage of parents encouraging junk food as a reward was significantly lower than the pretest ($P<.001$). Most parents (41/56, 73%) agreed that the computer game can improve their children's nutritional knowledge, reduce the frequency of junk food intake, and improve dietary behavior. Importantly, the intake frequencies of candy and chocolate as well as ice cream and ice pops decreased in the experimental group after the computer game intervention. The top 3 junk foods among the participating children were biscuits, candy and chocolates, and juice and sweetened beverages (data not shown).

XSL•FO
RenderX

Table 1. Descriptive characteristics of participants.

Variable and item	All (n=104)	Experimental group (n=56)	Control group (n=48)	*P* value
Age (years), mean (SD)	5.14 (0.27)[a]	5.13 (0.31)	5.15 (0.23)	.67
Gender, n (%)				.38
Male	59 (56.7)	34 (60.7)	25 (52.1)	
Female	45 (43.3)	22 (39.3)	23 (47.9)	
Birth ranking, n (%)				.33
First	43 (41.3)	22 (39.3)	21 (43.8)	
Second	50 (48.1)	30 (53.6)	20 (41.7)	
Third	11 (10.6)	4 (7.1)	7 (14.6)	
Educational degree of the father, n/N (%)[a]				.32
High school or below	46/102 (45.1)	29/55 (52.7)	17/47 (36.2)	
College	50/102 (49.0)	23/55 (41.8)	27/47 (57.4)	
Graduate school or above	6/102 (5.9)	3/55 (5.5)	3/47 (6.4)	
Educational degree of the mother, n/N (%)[a]				.14
High school or below	39/103 (37.9)[a]	24/55 (43.6)	15/48 (31.3)	
College	58/103 (56.3)	30/55 (54.5)	28/48 (58.3)	
Graduate school or above	6/103 (5.8)	1/55 (1.8)	5/48 (10.4)	

[a]Some participants are single parents.

Table 2. The computer game scores of the preschool children.

Variable	Score	*t* value (*df*)	*P* value
First week, mean (SD)	25 (23)	N/A[a]	N/A
Second week, mean (SD)	30 (29)	–1.11 (55)	.27
Third week, mean (SD)	38 (33)[b]	–2.4 (55)	.02
Fourth week, mean (SD)	37 (25)[c]	–2.98 (55)	<.001

[a]N/A: not applicable.

[b]Third week > first week.

[c]Fourth week > first week.

Table 3. The scores of nutritional knowledge.

Variable	All	Experimental group (n=56)	Control group (n=48)	*P* value
Pretest, mean (SD)	80 (15)	81 (16)	78 (14)	.26
Posttest, mean (SD)	85 (14)	89 (10)	80 (15)	<.001
Posttest-pretest, mean (SD)	5 (12)	8 (12)	1 (10)	.002

XSL•FO
RenderX

Table 4. Frequency of junk food intake between the experimental (n=56) and control (n=48) groups.

Variable and group	<1 time/month, n (%)	1-3 times/month, n (%)	>1 time/week, n (%)	*P* value
Pretest				
Candy and chocolate				.50
Experimental	16 (29)	19 (34)	21 (38)	
Control	9 (19)	19 (40)	20 (42)	
Ice cream and ice pops				.48
Experimental	22 (39)	25 (45)	9 (16)	
Control	14 (29)	23 (48)	11 (23)	
Cake				.60
Experimental	33 (59)	21 (38)	2 (4)	
Control	24 (50)	21 (44)	3 (6)	
Biscuit				.72
Experimental	6 (11)	21 (38)	29 (52)	
Control	3 (6)	19 (40)	26 (54)	
Soft drinks				.70
Experimental	36 (64)	15 (27)	5 (9)	
Control	27 (56)	16 (33)	5 (10)	
Fruit juice and sugary drinks				.02[a]
Experimental	18 (32)	18 (32)	20 (36)	
Control	12 (25)	28 (58)	8 (17)	
Posttest				
Candy and chocolate				.54
Experimental	20 (36)	21 (38)	15 (27)	
Control	13 (27)	18 (38)	17 (35)	
Ice cream and ice pops				.21
Experimental	31 (55)	21 (38)	4 (7)	
Control	20 (42)	20 (42)	8 (17)	
Cake				.92
Experimental	36 (64)	17 (30)	5 (9)	
Control	29 (60)	16 (33)	3 (6)	
Biscuit				.98
Experimental	7 (13)	20 (36)	29 (52)	
Control	6 (13)	18 (38)	24 (50)	
Soft drinks				.52
Experimental	36 (64)	13 (23)	7 (13)	
Control	27 (56)	16 (33)	5 (10)	
Fruit juice and sugary drinks				.31
Experimental	22 (39)	22 (39)	12 (21)	
Control	15 (31)	26 (54)	7 (15)	

[a]*P*<.05.

XSL•FO
RenderX

Table 5. The relationship between computer game intervention and nutritional knowledge score.

Variable	Coefficients	SE	95% CI	*P* value[a]
Nutritional knowledge score[b]	8.45	2.38	3.72-13.19	.001

[a]Adjustment for age, gender, parents education level, and parents' occupation in the linear regression.

[b]Posttest-pretest score.

Discussion

Principal Findings

Our results show that children's nutritional knowledge scores in the posttest were significantly higher than those in the pretest within the experimental group, whereas the scores of the control group were not significantly different between the pre- and posttest. This shows that Healthy Rat King, an educational game, improves the nutritional knowledge of children. This result is consistent with findings from other studies that computer games could help children correctly distinguish between healthy and unhealthy foods, while also improving the nutritional knowledge of children [9,11,14,15]. The sound, light, audio, and visual effects of computer games could arouse learning motivation and interest among children. By integrating computer games into education, children can maintain their enthusiasm to learn. Computer games have been one of the important methods to arouse active learning motivation and encourage enthusiasm among learners [16].

There was no significant difference in junk food intake between the experimental and control groups in the posttest. However, among children in the experimental group, intake frequencies of both chocolate and ice cream and ice pops decreased after the intervention. Children preferred to eat healthy foods after playing the computer game, Healthy Rat King, suggesting positive changes in attitudes and behaviors. This result is similar to that of previous studies in which school-aged children who received a computer game education increased their consumption of vegetables, fruits, and other healthy foods, and reduced junk foods such as French fries, candy, chocolate, cakes, soft drinks, fried foods, and sugary drinks, compared with children who did not receive the computer game intervention [9-11,17,18]. The results of these studies clearly indicate that the application of computer games to nutrition education can effectively improve the eating habits and promote healthy behaviors among young children.

In the experimental group, the percentage of children who refused to consume junk food was significantly higher than that in the pretest, while parents' use of junk food as a reward for children also decreased after the computer game intervention; however, there was no significant difference in the control group. This result shows that parents have made also some changes (reduction in the provision of junk food as a reward) in their children's diet based on their children's opinions (reject junk food). Young children's eating behavior is affected by many factors, including physiological, experience, learning, and social factors. Among these factors, parent's eating behavior and family dietary habits are the key to cultivating good eating habits and behaviors among preschool children. If the parents have a high intake of junk food or frequently offer junk foods to their children, then their children will also have these poor eating habits. Therefore, parental eating habits or family food provision has a significant influence on preschool children's nutritional knowledge and eating behavior [11,19]. Although intake frequencies of candy and chocolate as well as ice cream and popsicles have decreased following the computer game intervention in the experimental group, there was no significant difference in junk food eating behavior between the groups in the posttest. To the best of our knowledge, the family's eating habits directly impact those of children, and so they need a long time to develop their own eating habits. However, the computer game intervention period in this study lasted for only 4 weeks, which may explain the absence of significant difference in the junk food intake frequencies between these 2 groups in the posttest.

Studies have suggested that computer games can improve children's nutritional knowledge attitudes and behaviors. However, computer games have some inherent problems. Playing computer games for a long time could easily cause eye fatigue, which leads to a decreased vision. Early childhood is an important stage for eye development, so children should avoid eye damage due to use the excessive use of computer for a long time. Although computer games enrich learning, it could also increase certain injuries [20]. For example, computer games may separate young children from useful gaming experiences (physical activity), reduce opportunities for developing social skills, and cause problems such as lack of social interaction. In addition, unhealthy types of computer games or game plots might be harmful to young children's physical and mental development [20].

The results of this study show that the application of computer games in nutrition education for preschool children can effectively promote nutritional knowledge. This study can provide inspiration for the design of computer games and teaching tools suitable for preschool children.

Limitations

The computer game, Healthy Rat King, designed for the purpose of nutrition education is simple, cute, interesting, and fun. Although the game meets the psychological needs of children, some children face difficulties in using computers, which might lead to a bias in the computer game scores. Computer game designers may also be ignorant about the development status of children, which might complicate the gaming operations for children, and it might be beyond the abilities of some children. As motor development of preschool children is not mature, somatosensory games could be integrated into computer games in the future.

In addition, the computer game play uses keyboard and mouse. Because the physical and cognitive development of preschool

XSL•FO
RenderX

children is immature at this age, maneuvering and using keyboard and mouse controls were not easy. Therefore, during the game, researchers can use stickers of different colors to be pasted on the keyboard or mouse, which is convenient for children to identify and operate the game.

In this study, the 4-week computer game intervention had a short-term effect on improving children's nutritional knowledge and behavior changes. Another study also showed that short-term computer game intervention can increase nutritional knowledge; however, it might not be enough to greatly increase children's nutritional knowledge and actual eating behavior in the long term [9]. Hence, a follow-up study will be performed to investigate which duration of the computer game intervention would change the long-term eating behaviors of children in the future. Moreover, the future research requires a larger sample size to understand the effectiveness of computer games in nutritional and health education for young children.

Conclusion

Teaching using a computer game could improve children's nutritional knowledge. However, the intake frequency of junk food among children in the experimental group showed no significant difference from those in the control group in the posttest.

Acknowledgments

We appreciate all the participating preschools and principal for their support and cooperation.

Authors' Contributions

All authors have participated in the work including design of the study, data collection and analysis, interpretation of data, writing and revision of the manuscript, and final approval of the version to be submitted and published.

Conflicts of Interest

None declared.

References

1. Bhushan C, Taneja S, Khurana A. Burden of Packaged Food on Schoolchildren: Based on the CSE survey ?Know Your Diet? 2017. Centre for Science and Environment, New Delhi. URL: http://cdn.cseindia.org/attachments/0.37055100_1520598445_burden-packaged-food-school-children.pdf [accessed 2021-07-02]
2. Survey report on children's snack consumption in 2013. Child Welfare Alliance. URL: https://www.children.org.tw/research/detail/70/610 [accessed 2021-06-20]
3. Wiles NJ, Northstone K, Emmett P, Lewis G. 'Junk food' diet and childhood behavioural problems: results from the ALSPAC cohort. Eur J Clin Nutr 2009 Apr 05;63(4):491-498 [FREE Full text] [doi: 10.1038/sj.ejcn.1602967] [Medline: 18059416]
4. Maluccio J, Hoddinott J, Behrman J, Martorell R, Quisumbing A, Stein A. The impact of improving nutrition during early childhood education among guatemalan adults. The Economic Journal 2009;l19:734-763. [doi: 10.1111/j.1468-0297.2009.02220.x]
5. Payab M, Kelishadi R, Qorbani M, Motlagh ME, Ranjbar SH, Ardalan G, et al. Association of junk food consumption with high blood pressure and obesity in Iranian children and adolescents: the CASPIAN-IV Study. J Pediatr (Rio J) 2015;91(2):196-205 [FREE Full text] [doi: 10.1016/j.jped.2014.07.006] [Medline: 25449791]
6. Feng J, Wang WJ. Discussion on modern nutrition education model in kindergarten. E-magazine for the new education era 2018;31:11-35.
7. Vandewater EA, Rideout VJ, Wartella EA, Huang X, Lee JH, Shim MS. Digital childhood: electronic media and technology use among infants, toddlers, and preschoolers. Pediatrics 2007 May;119(5):e1006-e1015. [doi: 10.1542/peds.2006-1804] [Medline: 17473074]
8. McKenney S, Voogt J. Technology and young children: How 4–7 year olds perceive their own use of computers. Computers in Human Behavior 2010 Jul;26(4):656-664. [doi: 10.1016/j.chb.2010.01.002]
9. Hermans RC, van den Broek N, Nederkoorn C, Otten R, Ruiter EL, Johnson-Glenberg MC. Feed the Alien! The Effects of a Nutrition Instruction Game on Children's Nutritional Knowledge and Food Intake. Games Health J 2018 Jun;7(3):164-174. [doi: 10.1089/g4h.2017.0055] [Medline: 29634366]
10. Sharma SV, Shegog R, Chow J, Finley C, Pomeroy M, Smith C, et al. Effects of the Quest to Lava Mountain Computer Game on Dietary and Physical Activity Behaviors of Elementary School Children: A Pilot Group-Randomized Controlled Trial. J Acad Nutr Diet 2015 Aug;115(8):1260-1271. [doi: 10.1016/j.jand.2015.02.022] [Medline: 25912520]
11. Espinosa-Curiel IE, Pozas-Bogarin EE, Lozano-Salas JL, Martínez-Miranda J, Delgado-Pérez EE, Estrada-Zamarron LS. Nutritional Education and Promotion of Healthy Eating Behaviors Among Mexican Children Through Video Games: Design and Pilot Test of FoodRateMaster. JMIR Serious Games 2020 Apr 13;8(2):e16431 [FREE Full text] [doi: 10.2196/16431] [Medline: 32281539]

XSL•FO
RenderX

12. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for Health for Children-Current Status and Needed Research. Games Health J 2016 Feb;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
13. Exemption from Review | Research Ethics Center of China Medical University and Affiliated Hospitals. Center of China Medical University and Affiliated Hospitals. URL: https://rrec.cmu.edu.tw/?q=zh-hant/node/30 [accessed 2021-07-02]
14. Schneider KL, Ferrara J, Lance B, Karetas A, Druker S, Panza E, et al. Acceptability of an Online Health Videogame to Improve Diet and Physical Activity in Elementary School Students: "Fitter Critters". Games Health J 2012 Aug;1(4):262-268 [FREE Full text] [doi: 10.1089/g4h.2012.0009] [Medline: 24761317]
15. Turnin MC, Tauber MT, Couvaras O, Jouret B, Bolzonella C, Bourgeois O, et al. Evaluation of microcomputer nutritional teaching games in 1,876 children at school. Diabetes Metab 2001 Sep;27(4 Pt 1):459-464. [Medline: 11547219]
16. Yang L. Let students walk into computer games in a healthy way-Investigation and analysis of the current situation of students playing computer games in a certain elementary school in Nanjing. Curriculum Education Research 2013;7:192-193.
17. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
18. Majumdar D, Koch PA, Lee H, Contento IR, Islas-Ramos ADL, Fu D. "Creature-101": A Serious Game to Promote Energy Balance-Related Behaviors Among Middle School Adolescents. Games Health J 2013 Oct;2(5):280-290 [FREE Full text] [doi: 10.1089/g4h.2013.0045] [Medline: 24761326]
19. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors Influencing Children's Eating Behaviours. Nutrients 2018 May 31;10(6):706 [FREE Full text] [doi: 10.3390/nu10060706] [Medline: 29857549]
20. Lee HC, Hsieh IY. Analyzing the influence that play computer game in the early childhood and its implications on education. Journal of Preschool Education Research 2007;1:101-105 [FREE Full text]

Edited by N Zary; submitted 25.08.21; peer-reviewed by E Hohman, Q Zhou; comments to author 11.12.21; revised version received 19.01.22; accepted 28.04.22; published 01.07.22.

Please cite as:
Chang IC, Yang CY, Yen CE
The Effects of a Computer Game (Healthy Rat King) on Preschool Children's Nutritional Knowledge and Junk Food Intake Behavior: Nonrandomized Controlled Trial
JMIR Serious Games 2022;10(3):e33137
URL: https://games.jmir.org/2022/3/e33137
doi:10.2196/33137
PMID:35776502

©Ing-Chau Chang, Cheng-Ying Yang, Chin-En Yen. Originally published in JMIR Serious Games (https://games.jmir.org), 01.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Preliminary User Evaluation of a New Dental Technology Virtual Simulation System: Development and Validation Study

Mengwei Pang[1,2,3,4*], BDS; Xiaohan Zhao[5*], MD; Daiyu Lu[1,2,3,4], BDS; Yihan Dong[1,2,3,4], BDS; Lin Jiang[1,2,3,4], DDS, PhD; Jie Li[1,2,3,4], PhD; Ping Ji[1,2,3,4], DDS, PhD

[1]College of Stomatology, Chongqing Medical University, Chongqing, China

[2]Chongqing Key Laboratory of Oral Diseases and Biomedical Sciences, Chongqing, China

[3]Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education, Chongqing, China

[4]Stomatological Hospital of Chongqing Medical University, Chongqing, China

[5]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Ping Ji, DDS, PhD
College of Stomatology, Chongqing Medical University
426# Songshibei Road, Yubei District
Chongqing, 401147
China
Phone: 86 023 8886 0026
Fax: 86 023 8886 0222
Email: jiping@hospital.cqmu.edu.cn

Abstract

Background: With the advancements in the dental health care industry, the demand for dental technicians has increased. Dental technicians should be thoroughly assessed and trained in practical skills and pass professional certification examinations to ensure that they are competent to work closely with dentists. Unfortunately, such training courses and tests are in short supply worldwide. The use of virtual simulation technology can help solve these problems.

Objective: This study presents a new strategic framework design for a certified dental technician practical examination called as the certified Objective Manipulative Skill Examination of Dental Technicians (OMEDT), which is based on the Objective Structured Clinical Examination (OSCE). We present the development and validation of the OMEDT system, a new virtual simulated training system, to meet the demands of the OMEDT framework. The combination of OMEDT and the OMEDT system can solve the complex problems encountered in the certified dental technician practical examination with excellent efficiency, high quality, and low cost.

Methods: The OMEDT framework design was constructed according to the OSCE guide and the Chinese vocational skill standards for dental technicians. To develop the OMEDT system, we organized a new framework based on the virtual learning network platform, the haptic feedback system, and the real-time dental training and evaluation system. The effectiveness evaluation of the OMEDT system was divided into 2 phases: in the first phase, 36 students were recruited to use the test module to finish the task and their performance data were collected and analyzed; and in the second phase, a questionnaire was administered to 30 students who used the system for their studies and graduation exams.

Results: The OMEDT and the corresponding skill training virtual simulation OMEDT system were developed, and preliminary user evaluation was performed to assess their effectiveness and usefulness. The OMEDT system was found to improve students' practical skills by training with the evaluation results. In addition, several key research topics were explored, including the effects of positive feedback of the knowledge of results on the improvement of the students' skill level and the common sense transformation of educators in the virtual simulation technology environment.

Conclusions: The development of OMEDT and the OMEDT system has been completed and their effectiveness has been verified.

(JMIR Serious Games 2022;10(3):e36079) doi:10.2196/36079

XSL•FO
RenderX

KEYWORDS

virtual simulation; dental technology; OSCE; virtual reality; dentistry; dental technician; framework; certified dental technician; development; validation; serious game; dental technology; dental; technology

Introduction

The dental health care industry is a technically oriented discipline [1]. With the advancements in the dental industry, the demand for dental technicians has increased. Dental technology, which has originated from traditional dentistry, has become an independent field of study in China where students are trained in the latest technology of dental prostheses [2]. Skill is critical for the professional development of dental technicians, including the cognitive ability to apply theoretical concepts, professional psychomotor skills in dental technology laboratories, affective skills of communicating with dentists, and the responsibility of maintaining the prostheses quality. The skill level of a dental technician determines the quality of the denture processing, which then affects the success of the treatment. Unfortunately, relevant skill training courses incur high costs, as they require a significant capital investment in equipment and materials and high educator/student ratios to maintain close supervision [3-7]. Moreover, limitations in the development of dental technology are reflected by the lack of a robust certification system in the industry [8,9]. The certified dental technician (CDT) practical examination, which is the vocational assessment of dental technology education, is not conducted worldwide.

In light of the above limitations, the assessment of dental technology skills can be developed following the pattern of the stomatology professional certification practical examination. The Objective Structured Clinical Examination (OSCE) is a widely used strategy for the stomatology professional certification practical examination worldwide. Dental institutes in various countries, including China, use OSCEs for formative and summative assessments of students' clinical competence [10-12]. OSCE is a tool used for evaluating performance in simulated environments; it consists of a series of standardized assessment stations, requiring students to perform several tasks based on clinical situations. Either standardized or alternative simulators are used. The examinees are assessed using the same stations and rating schemes within the same timeframes, and the examiners assess the examinees' application of knowledge to practice [13]. Over the past few decades, OSCE has proven to be a valid and reliable tool for assessing all 3 learning domains of cognitive skills, psychomotor skills, and affective skills [11,14]. However, for dental technology students, OSCE is a complex, resource-intensive, and time-intensive assessment examination. Unlike dental clinical skills, skills associated with fabricating corrective devices and replacements require more process time, resources, semifinished products, etc. Thus, it is necessary to embrace new technologies, reset the strategic framework, create more immersive simulation environments, and transform OSCEs at the application level to meet the needs of CDT practice examinations.

Recent research studies [15,16] have found that virtual reality simulation technologies could create more immersive simulation environments and may positively affect dental education. The existing virtual simulation technology is mainly divided into the virtual learning network platform (VLNP), the haptic feedback system, and the real-time dental training and evaluation system (RDTES). VLNP is usually designed as a series of web-based skill training modules. It helps students grasp the operational essentials and improves cognitive ability and affective skills. Compared with virtual reality technology, VLNP is less realistic, but it has the advantage of being less expensive and can be built on serious game theories [15,16]. For developing a multimodule virtual simulation system, VLNP is a more suitable option considering the costs. Inexperienced dental technicians or students may be exposed to occupational hazards and risks, thereby threating public safety. Such hazards can be avoided by using haptic feedback technology to build virtual training environments. Haptic interaction has been successfully applied in dental technology skill training to simulate specific tasks [17]. RDTES is a preclinical simulator that provides real-time image processing using 3D graphics and video recordings. Recent studies [18,19] have shown that if students see illustrations of their procedures, they can understand any inadequacies in their skills objectively and visually. However, another study [20] has shown that RDTES cannot wholly replace the feedback that educators give students during teaching activities. In RDTES technology, the inclusion and recording of educators' assessments have always been an ongoing problem.

The advantages of virtual simulation technologies include safety, time saving, economic benefits, ethical benefits, increased frequency and relevance of training, and teaching error management [21-38]. The disadvantages include less realistic force feedback simulation, insufficient training content, and unquantified evaluation of the training results [39]. To meet the requirements of the CDT practical examination, many technical and structural issues need to be addressed. Unfortunately, although many attempts have been made to develop a dental simulator by different manufacturers for the psychomotor skills training of dental students, there has never been a system designed explicitly for the professional performance training of dental technology students and the CDT practical examination. Thus, this study presents a new strategic framework design for the CDT practical examination based on OSCE, called as the certified Objective Manipulative Skill Examination of Dental Technicians (OMEDT). Then, we present the development and validation of a new psychomotor skill training virtual simulation system, the OMEDT system, to meet the OMEDT framework's demands. The OMEDT system is based on the VLNP–haptic feedback–RDTES (V-H-R) architecture, specifically for dental technology students and the CDT practical examination. Experiments were designed to evaluate the effectiveness of this system. The purpose of this study was to develop a new examination framework and the corresponding skill training virtual simulation system for dental technology students, evaluate the effect of the trainings through virtual simulation on dental technology students' professional

XSL•FO
RenderX

performance during preclinical laboratory work, and discuss the possibility of utilizing virtual simulation technology in the CDT practical examination.

Methods

Framework Design of OMEDT

The OMEDT framework design was constructed in strict accordance with the OSCE:AMEE Guide [11,12] and the national vocational skill standards for dental technicians [40], which were presented by the National Health Commission of the People's Republic of China. Based on the talent cultivation mode oriented by industry demand, most universities follow this standard in designing professional performance training strategies and curricula for their dental technology majors. This approach ensures that most universities' dental technology graduates have the ability to pass the OMEDT and that graduates who pass the OMEDT are prepared to work in the industry.

For critical technologies with common characteristics in different technological processes, the design steps in this strategy have been deleted, combined, and supplemented. Moreover, as per the time taken to complete the technical processes, the examination is divided into short-term completion projects and long-term completion projects and the individual combination of the examination stations is determined by random draws. Based on the development concept of "combining virtual with the actual environment," the OMEDT system must be used to complete the virtual simulation site in the OMEDT.

Design and Compilation of the OMEDT System

The OMEDT system was developed considering the following requirements:

1. Increase the opportunities of dental technology students' professional psychomotor skills training.
2. Turn complex experiments into serious games to gradually acquire new problem-solving skills and domain knowledge.
3. Decrease the requirement of dental laboratory equipment and the waste of experimental materials due to inexperienced psychomotor skills.
4. Spend more time learning skills and less time waiting in line for faculty feedback, shortening the amount of time spent in the dental laboratory.
5. Establish an instructor–student technician communication mechanism that provides immediate feedback.
6. Analyze and store students' training data and instructors' comment data.
7. Cover the major process of dental technology's curriculum and satisfy the demand of the CDT practical examination.

Based on the above considerations, the OMEDT system selected the following contents for module development in the first phase of the study: (1) prosthesis casting process simulation, (2) tooth morphology sketch design simulation, (3) removable partial denture (RPD) design (D-RPD) simulation, (4) esthetic dental photography and design simulation, (5) dies diagnostic simulation, and (6) tooth engraving haptic feedback simulation. The structure of the data management module is designed to manage the above contents. The OMEDT system was developed with the Unity engine (Unity Technologies, Unity Pro) and the C++11 and C# 4.0 developing languages. The teeth data were collected by computed tomography and intraoral 3D scanning. AutoCAD 3ds Max was used to refine the shape of the geometric models. To achieve vivid physical simulation, the physical skeleton of the geometric environment was constructed. The gypsum blocks used for carving were modeled as triangle meshes. With the 6 degrees of freedom (6-DOF) haptic rendering software development toolkit provided by the State Key Laboratory of Virtual Reality Technology and Systems of Beihang University, we succeeded in simulating the operation process of tooth carving with force feedback [41,42].

Performance Evaluation of the OMEDT System

The performance evaluation of the OMEDT system was divided into 2 phases. Phase I consisted of user tests for specific modules. Phase II consisted of a questionnaire survey for long-term users.

Phase I

The OMEDT was initially designed as a whole life cycle training framework for promoting students' professional performance. It consists of many different modules, including casting and D-RPD. A single student should undertake specific modules to train according to his or her learning progress. For the construction validity evaluation of the OMEDT system, the V-H-R architecture and the learning structure of students should be considered simultaneously. Test modules with high coverage in VLNP and RDTES dimensions should be selected to perform experimental evaluation. Therefore, the D-RPD module was selected to perform the evaluation experiment. D-RPD is an important core content in dental technology majors, and the courses related to this content will last at least 3 semesters in the training plan, with wide coverage in the learning period. The data volume and data dimension in the D-RPD module are relatively full, which is a great challenge for the execution of the data management module. Experiments using this module can check the execution effect of VLNP and RDTES in the V-H-R architecture at the same time. D-RPD is the most difficult core course for students of stomatology technology to fully master because of its complex content and the need for many instructors with extensive experience in D-RPD. The construction validity of the OMEDT can be fully illustrated by the verification of the difficulty levels in the modules.

In the first phase of the short-term system usability tests, 36 students majoring in dental technology were recruited to participate in the experiments, and they signed informed consent forms. To ensure the consistent level of expertise, all participants were from the same class at the College of Stomatology, Chongqing Medical University and received the same level of instruction in D-RPD theory. Before formal evaluation experiments, every participant received sufficient pretraining to become familiar with the OMEDT system and the testing module, including the use of computer equipment, software-related operations, etc. The completion of the pretraining is based on students being able to draw RPD designs by using various tools within the D-RPD module. Each student took 7 weeks to complete the formal evaluation experiments. The students were required to complete and submit the design of RPDs based on actual clinical cases with a D-RPD module

XSL•FO
RenderX

per week. Students used their own personal computers to complete tasks in different network environments. These tasks were designed as additional assignments for the course. Students were told that completing these tasks would not give them extra points but would help them improve their design skills. Each design was scored by 2 experienced dental technology instructors with the testing module. The experiment was monitored by the OMEDT system developer team, which consisted of 3 technical support engineers and 1 dental technology instructor to deal with any potential system bugs. Data were collected from 2 dimensions: task completion time and task score. The D-Manager module was used to register and extract the original data and download it as a table. The data were analyzed using GraphPad Prism 8.0.

Phase II

In the second phase, 1 questionnaire was administered to 30 dental technology students who used the OMEDT system for a 1-year study and graduate examination. The questionnaire was designed to collect user feedback on using the OMEDT system, including the system's efficiency improvement for learning and working, improvement of skills and theoretical knowledge, improvement of independent learning ability, and acceptance of using the system for examinations. This preliminary questionnaire was piloted using 10 students who did not participate in the main study to ensure the clarity of the questionnaire. On the basis of the feedback, a panel discussion was organized by 2 senior instructors to revise some ambiguous words that might cause misunderstanding, avoid 2-in-1 questions, and make the questions more concrete. Finally, 9 items were ascertained, among which the participants rated 8 items by using a 5-point Likert scale. An open-ended question was used to collect suggestions for the future development and limitations of the OMEDT systems in dental technology education. The data were analyzed using GraphPad Prism 8.0.

Ethics Approval

The research ethics committee of the Affiliated Hospital of Stomatology, Chongqing Medical University approved this study (COHS-REC-2022 [LSNo 080]).

Results

OMEDT Framework

The OMEDT is organized as a 1-day examination, divided into a half-day short assessment and a half-day long assessment. The short assessment consists of 4 compulsory stations and 2 random stations. The 4 compulsory stations are set to take 15 minutes to complete, the 2 random stations are set to take 45 minutes to complete, and the candidate transfer time between the stations is considered as 5 minutes. The long assessment consists of 1 random test station, which is set to take 3 hours to complete. Since the total time taken to complete the short and long assessments are almost the same and the time in each station in the short assessment is fixed, the candidates need to randomly draw lots before the OMEDT begins to determine their test station combination and test sequence for long and short assessments.

OMEDT System

The OMEDT system supports high-quality real-time 3D interaction. Multiple art resources are built into the system, including pictures, 3D models, and animations. To guarantee the rendering quality, the number of polygon meshes of a single scene can be up to 10 million polygons, and the resolution is less than 0.01 mm. In addition, the update frequency can be as high as 120 per frame. Bezier curve drawing, spline curve drawing, and other drawing tools are supported in the OMEDT system; the minimum error of drawing is not higher than 0.01 mm. The tooth-carving part of the OMEDT system is based on the Unidental haptic-based dental simulator 6-DOF haptic rendering, and the simulation of bimanual haptic feedback is provided for the students to practice hand, eye, and foot coordination. Based on the V-H-R architecture, the OMEDT system has 7 main modules: D-Casting, D-Sketching, D-RPD, D-Smile, D-Dies, D-Carving, and D-Manager. These modules are for prosthesis casting (D-Casting), dental sketching (D-Sketching), D-RPD, esthetic dental photography and design (D-Smile), dies diagnosis (D-Dies), tooth carving haptic feedback (D-Carving), and integrated management (D-Manager). These are standardized into 4 main professional performance training design layouts:

1. Process simulation based on the demand of dental laboratory technology in reality and the regularization theory framework of the serious game carried out on the VLNP (D-Casting).
2. Design simulation based on the clinical cases and the blended learning instantaneously, carried out on the VLNP (D-Sketching, D-RPD, D-Smile).
3. Diagnosis simulation based on the clinical prescription and the modular assembly random database, carried out on the VLNP (D-Dies).
4. Haptic feedback simulation based on hand-eye-foot coordination and the assessment of multiple dimensions, carried out on the Unidental haptic device (D-Carving).

An integrated management system (D-Manager) based on data recording/integration/analysis was used to manage 6 modules. It can assist instructors in completing lecture teaching/homework review/objective structured examination design/test release and other functions while supporting students to practice freely and carry on the RDTES. The VLNP and RDTES modules are recommended to be accessed using the latest version of Google Chrome. The haptic feedback module was developed by Unidental, a third-generation dental simulator developed by the Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd (Figures 1-3).

XSL•FO
RenderX

Figure 1. Architectural design of the Objective Manipulative Skill Examination of Dental Technicians system. D-RPD: removable partial denture design; OMEDT: Objective Manipulative Skill Examination of Dental Technicians; RDTES: real-time dental training and evaluation system; VLNP: virtual learning network platform.

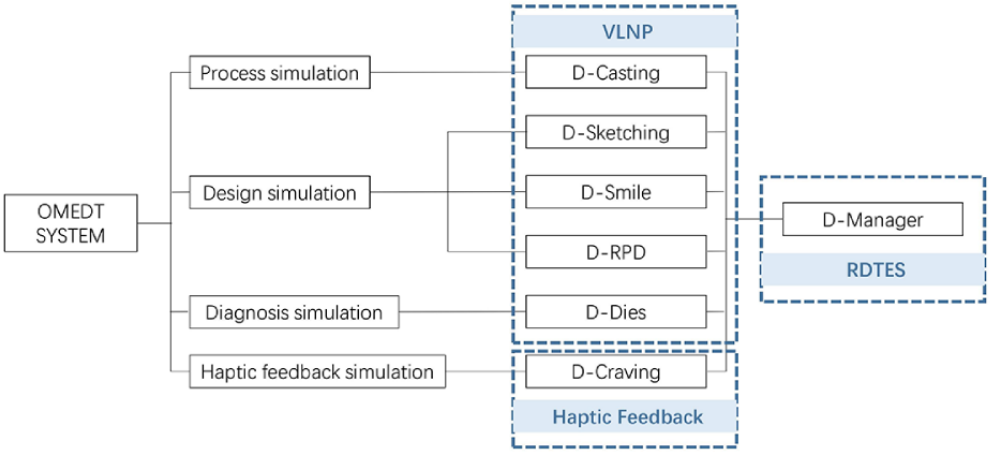


Figure 2. Architectural design of the Objective Manipulative Skill Examination of Dental Technicians and the corresponding modules in the Objective Manipulative Skill Examination of Dental Technicians system. CAD: computer-aided design; OMEDT: Objective Manipulative Skill Examination of Dental Technicians.

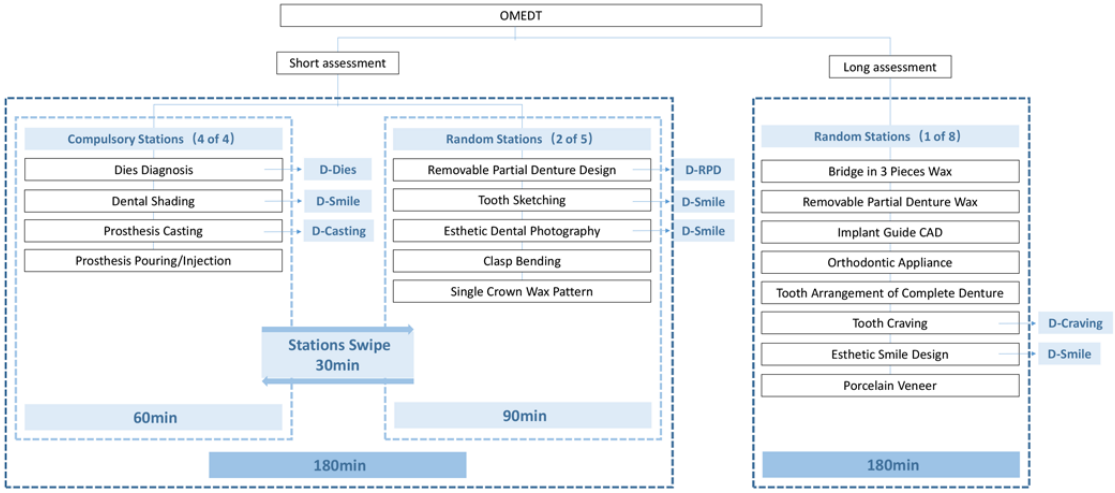
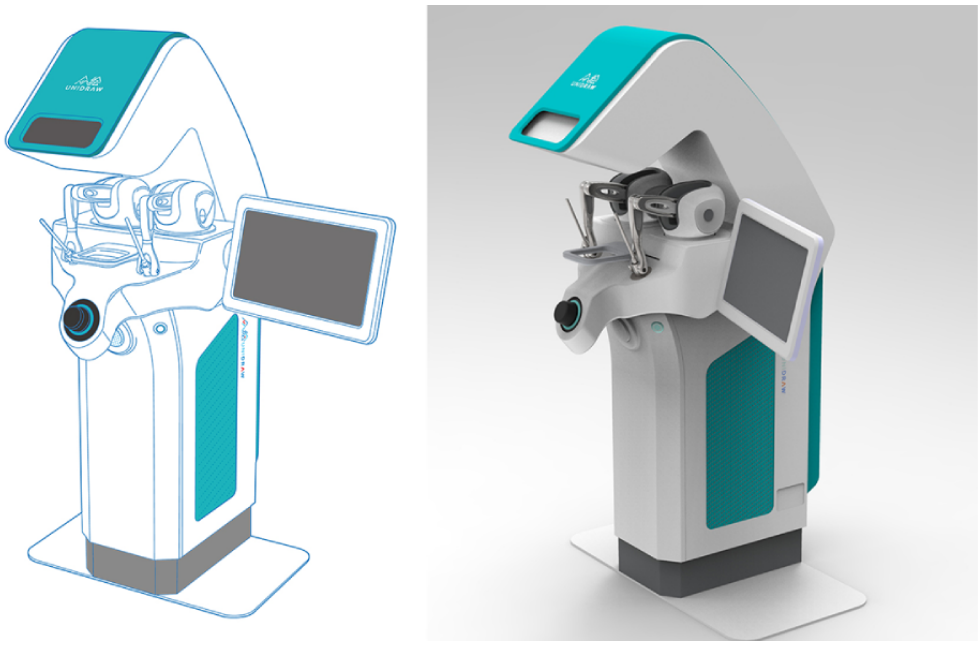


Figure 3. Unidental haptic feedback device mounted on the D-Carving module.

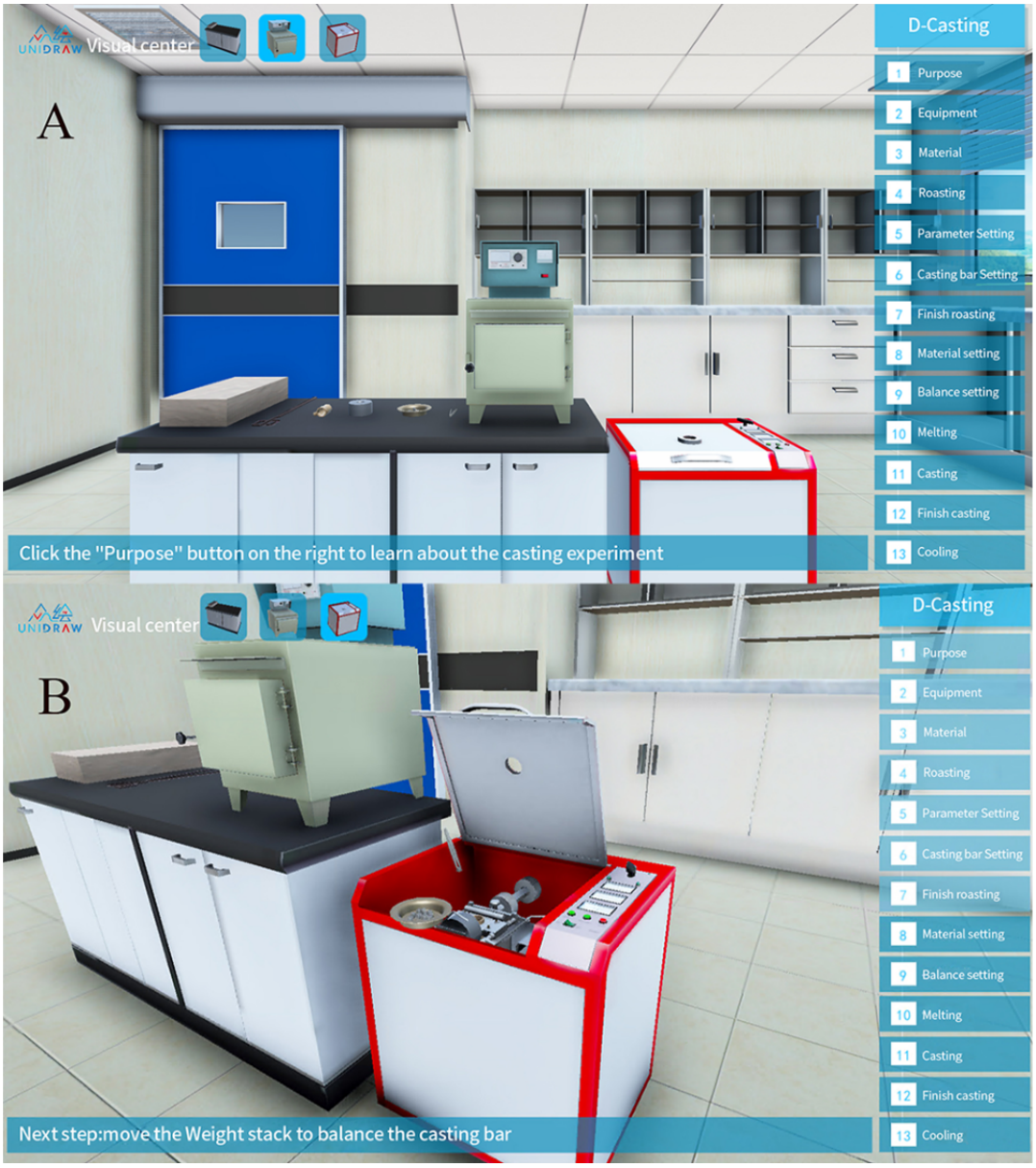


XSL•FO
RenderX

D-Casting

Unlike other virtual simulation systems or modules, D-Casting's approach to design follows the theoretical framework of serious games. In addition, the fabrication of a complete reward/punishment mechanism was designed to increase interest in the operation, and game technology is used to demonstrate those processes that could not be observed in reality, which leads to intense and passionate involvement, goals that motivate, and rules that provide structure. More importantly, D-Casting makes it possible to include the prosthesis casting process in the CDT practical examination, because lengthy processes can be speeded up in simulation, unaffected by semifinished products or any impact of human interference, thereby eliminating the potential for fraud and ethical issues (Figure 4).

Figure 4. Application programming interface of D-Casting. A: Panorama of the virtual casting laboratory in D-Casting; B: Prosthesis casting process simulation in D-Casting.

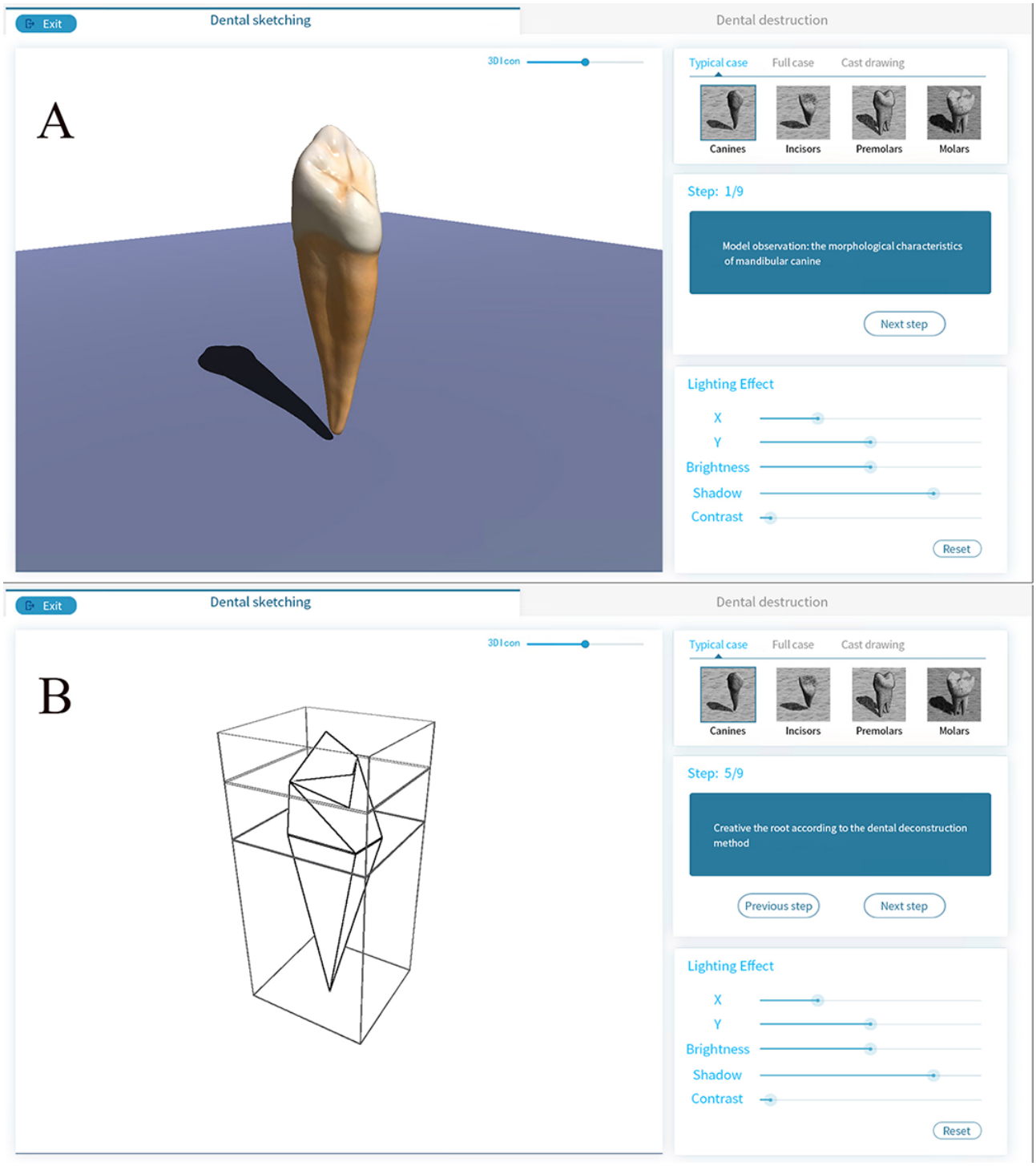


D-Sketching

Combined with an ingenious dental deconstruction method, D-Sketching provides a new approach to learning dental morphology sketches for inadequately trained dental technology students in the art area. The real-time rendering engine fabricates a movable virtual platform for students. They can observe the tooth with the simulation of the light, analyze the light and dark tones on teeth, and perform omnidirectional exercises on sketching to improve their skills. The "tooth assembly mode," based on dental deconstruction, is provided in D-Sketching. Students could use this module to finish the sketch without actual models based on mastering the morphology of the teeth. D-Sketching could help students build kinesthetic memories from sketching, review the dental anatomy from variable directions after rotation, and reconstruct by sketching. This process also conforms to Fitts and Posner's seminal model, the most well-used and well-recognized theory of psychomotor skill learning with a long history of use in health care [43] (Figure 5).

XSL•FO
RenderX

Figure 5. Application programming interface of D-Sketching. A: Panorama of the real-time rendering engine in D-Sketching; B: Sketching process based on the dental deconstruction method.

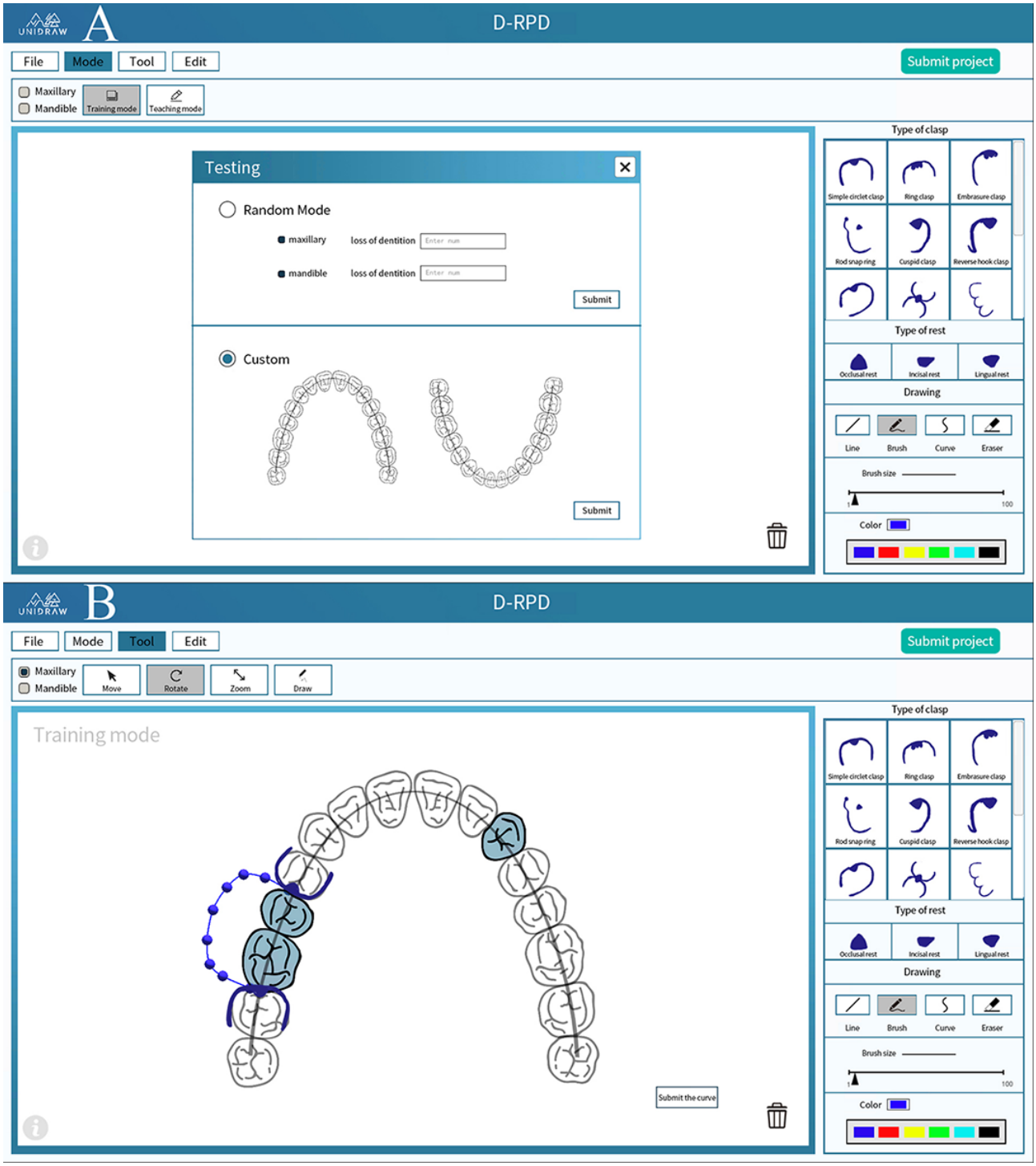


D-RPD

Based on the clinical case database, D-RPD provides overall solutions for students and instructors. The traditional mode of written D-RPD instructions was abandoned. The virtual dies reconstructed by the scanner were integrated into the database. After students observed the virtual model in D-RPD, they used D-RPD to fabricate electronic instructions. Instructors use D-RPD to review electronic instructions, and scores and feedback can be given in a timely manner. D-RPD supports students in performing independent exercises anytime and anywhere and provides the possibility for large-scale D-RPD examinations (Figure 6).

XSL•FO
RenderX

Figure 6. Application programming interface of the removable partial denture design. A: Free practice mode in the removable partial denture design module; B: Random testing mode in the removable partial denture design module.

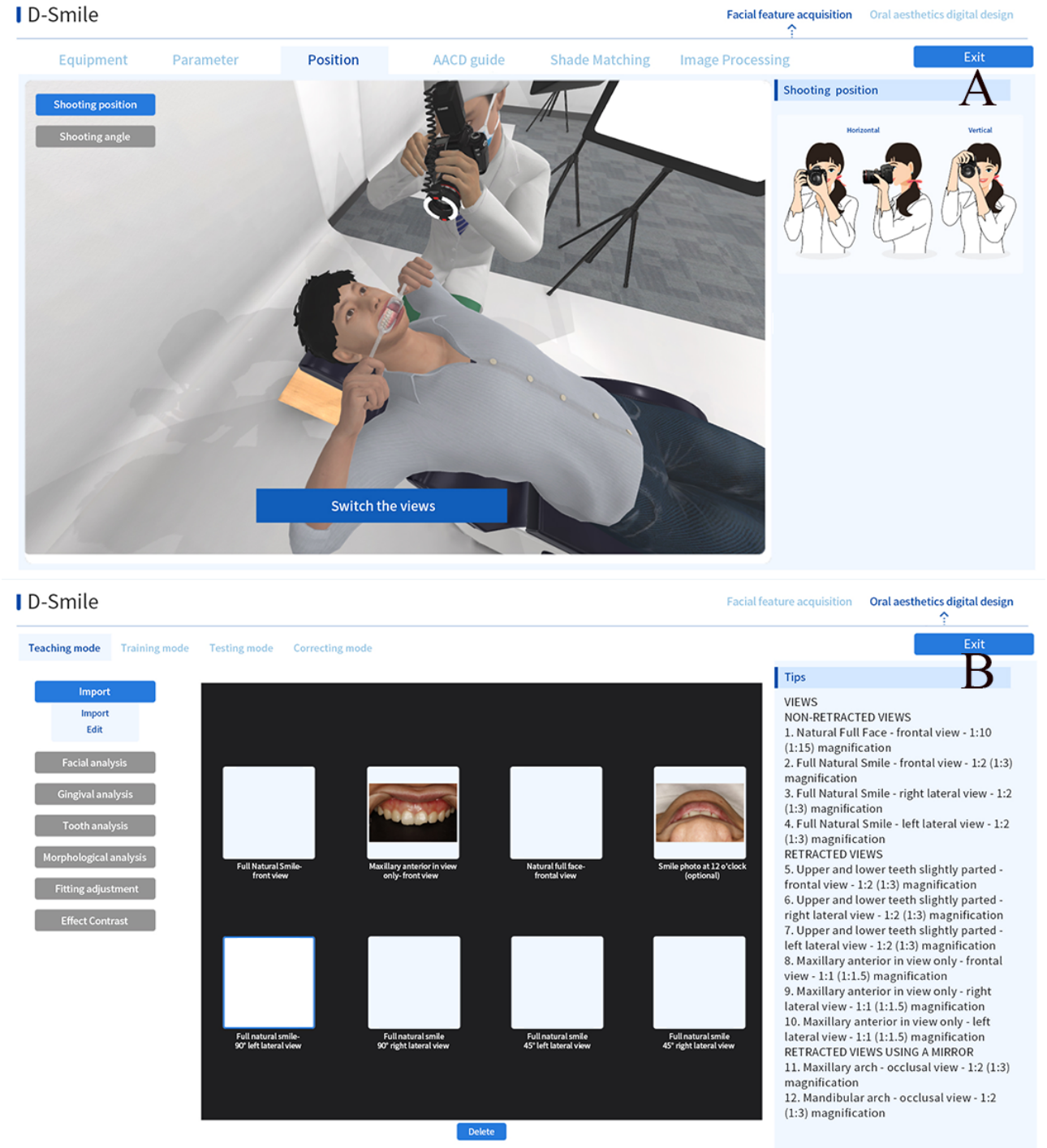


D-Smile

As per the American Academy of Cosmetic Dentistry [44], the whole process of dental macrophotography was simulated in D-Smile. The digital single lens reflex camera and the dental photography studio were disassembled in the simulation, and students could learn dental aesthetic photography techniques online without buying expensive equipment. Based on the teaching objectives, preoperative aesthetic design parts and clinical case databases were implanted into D-Smile. Similar to D-RPD, D-Smile supports students in performing independent exercises anytime and anywhere and provides the possibility for large-scale preoperative aesthetic design examinations (Figure 7).

XSL•FO
RenderX

Figure 7. Application programming interface of D-Smile. A: Panorama of a virtual oral photography studio in D-Smile; B: Oral digital aesthetic design based on facial features (referring to the accredited photography guide of the American Academy of Cosmetic Dentistry).



D-Dies

When the patient's intraoral gypsum dies enter the dental laboratory workflow, they usually need to be diagnosed by experienced dental technicians first, to ensure that the accuracy of the gypsum dies meets the production requirements, which also is the basis of the cognitive skills of dental technology students. Defective virtual models are classified strictly by defect types and randomly assembled in the database according to the logical framework of the modular assembly. In D-Dies, the defect dies reconstructed haphazardly will be diagnosed by the students, and then an electronic specification will be formed according to the suggestions. Instructors use D-Dies to review the electronic instructions, and scores and feedback can be given in a timely manner. Likewise, D-Dies supports students in performing independent exercises anytime and anywhere and provides the possibility for large-scale examinations (Figure 8).

XSL•FO
RenderX

Figure 8. Application programming interface of D-Dies. A: The virtual defect dies segmented by modularization in D-Dies; B: Randomly reconstructed virtual defect dies testing in D-Dies.



D-Carving

Owing to the tedious and complicated process of tooth carving and the limitations of labor and time, material cost, and safety, students have few opportunities to practice, and it is difficult for them to develop psychomotor skills. D-Carving is the first tooth carving training haptic system in the world. It uses a 6-DOF haptic rendering algorithm based on configuration optimization [42] to form stable multipoint interactions between the carving tool and the carving material within 1 millisecond. The system is equipped with different virtual tools such as dental ball drills, diamond needles, carving pens, and willow carving knives. D-Carving can simulate the texture of wax blocks, plaster, and other materials. The method of marching cube rendering based on a directed distance field and graphics processing unit acceleration can accurately simulate the dynamic change of the topological structure in material removal. The rendering accuracy reaches 0.1 mm to meet the training needs of fine-tooth carving. During the training, the standard control group can be added for comparison, and the evaluation module can be added (Figure 9).

XSL•FO
RenderX

Figure 9. Application programming interface of D-Carving. A: Panorama of the virtual carving steps in D-Carving; B: A student using the D-Carving module to practice tooth carving.



D-Manager

D-Manager can publish tests and assessments for any module in the background and support the instructor in marking homework online. Multiple dimensions of training data were collected, including students' training time in modules, students' work files, scores given by various instructors to the same student, etc. All data are stored and analyzed on the server to form charts. Instructors and students can master the training progress and adjust the training frequency and training time by checking the chart's data. Some VLNP parts of the OMEDT system can be used in iLab-x, which is the largest virtual simulation experiment online teaching center in China [45].

Experimental Results of System Effectiveness

Phase I

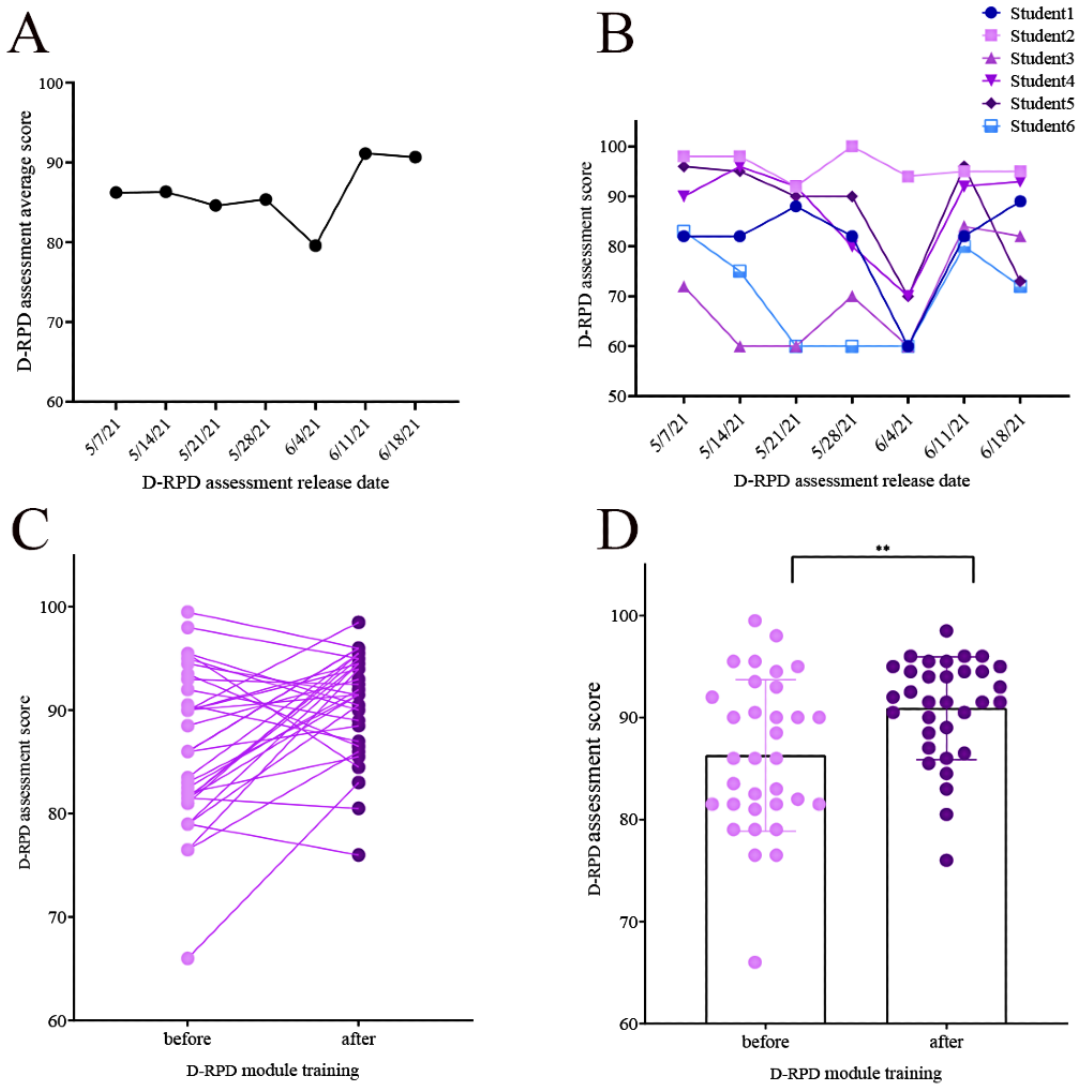
According to the design of performance evaluation, 33 students completed the 7-week practice task. Three students withdrew from the experiment for personal reasons; therefore, their scores were excluded from the statistical analysis. Owing to a system error in the D-Manager module in the first week, the task completion time data in the first week failed to be collected.

XSL•FO
RenderX

The resulting complete data set included the scores of the 33 students who had practiced D-RPD for 7 consecutive weeks, as well as task completion times for the next 6 weeks. Paired *t* test analysis was performed on the score-dimension data set, with the mean scores of the first 2 weeks as the pretest and the mean scores of the last 2 weeks as the posttest. Since the time consumption was not normally distributed, the time consumption data of the second week at the beginning of the test were used as the pretest, and the average time consumption data of the last 2 weeks were used as the posttest. Mann-Whitney analysis was performed on the time-dimension data set. The results of the complete data set analyses are shown in Figure 10 and Figure 11.

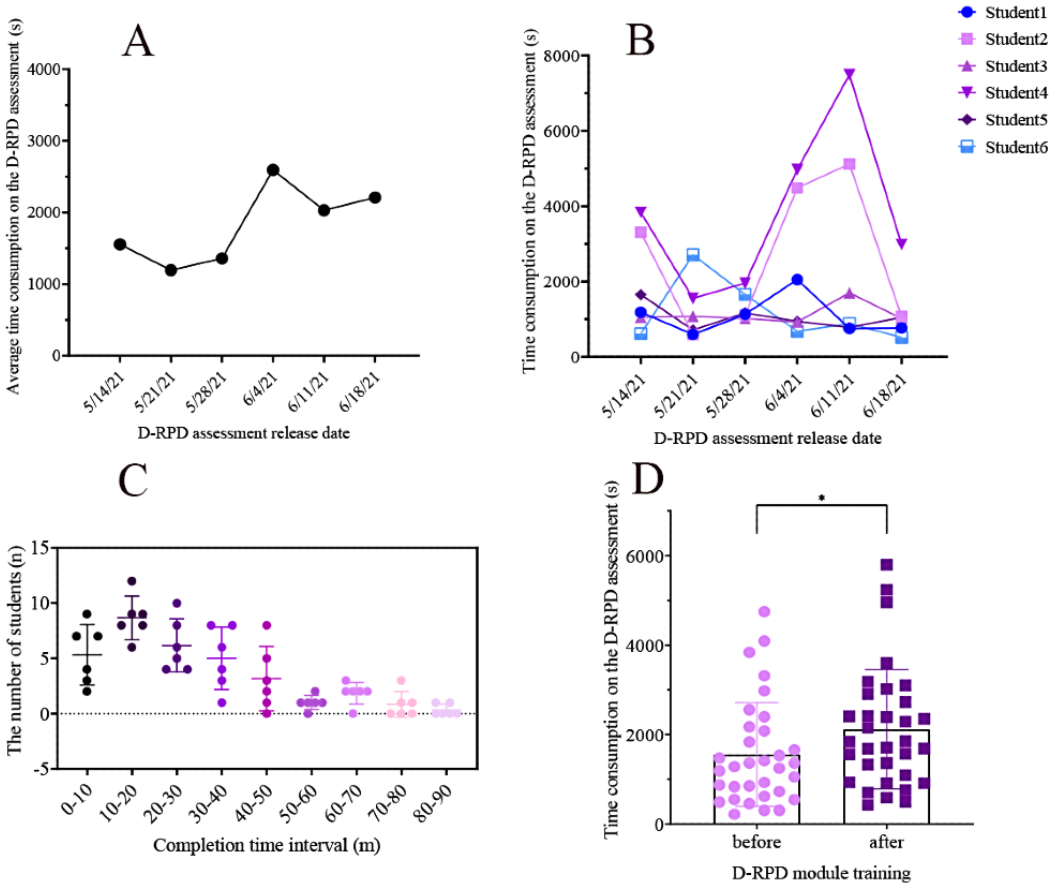
The results showed significant differences and fluctuations in the average scores. For the first 4 weeks, it was 85.65 out of 100. For the fifth week, it fell slightly to 79.60. At this stage, students reported some irritability such as feeling more stressed and less novelty in web-based work and more tough training exercises in the last 3 weeks. However, this finding was expected at the beginning of the design. Students' scores rebounded after they were given proper rest and educator incentives to continue with the exercise program. For the last 2 weeks, the average score rose above 91.15. The time consumption curve was highly complementary to the score curve. In the trough of the score curve, the time consumption curve reached its peak. The data of the 2 dimensions provide cross-evidence for educators to observe students' learning progress. At the same time, the data set supports hierarchical real-time observations from the individual to group to class dimensions and enables educators to identify areas of time consumption that were centrally distributed among the testers. Most participants in this experiment took 30-40 minutes to complete 1 D-RPD assessment. Statistical analysis of the score-dimension data set showed that students' D-RPD scores increased significantly after the D-RPD exercise (P=.002). The results of the statistical analysis of the time-dimension data set showed that the time consumed by the students on a single task increased significantly after the D-RPD exercise (P=.04).

Figure 10. Experimental results of the removable partial denture design (score dimension). A: Average time consumption of the removable partial denture design assessment (33 students in 7 weeks); B: Removable partial denture design assessment score of a sample team (6 students in 7 weeks); C and D: Paired *t* test analysis results of the score-dimension data set, P=.002. D-RPD: removable partial denture design.



XSL•FO
RenderX

Figure 11. Experimental results of the removable partial denture design (time dimension). A: Average time consumption of the removable partial denture design assessment (33 students in 6 weeks); B: Time consumption of the removable partial denture design assessment of a sample team (6 students in 6 weeks); C: Distribution of the removable partial denture design assessment completion (33 students in 6 weeks, mean [SD]); D: Paired Mann-Whitney analysis results of the time-dimension data set (P=.04). D-RPD: removable partial denture design.



Phase II

According to the design of performance evaluation, 28 of 30 students (93% response rate) completed the questionnaire and their responses were determined to be valid. In this study, Cronbach α was .980, suggesting that the reliability and the internal consistency of the statistics were sufficient. Obviously, all indicators exceeded the pass mark, proving the effectiveness of the system. Table 1 shows the mean (SD) score on each item.

In the open-ended questions, some students reported that systems were sometimes difficult to connect. However, investigations revealed that these problems were mainly due to errors, as users did not clear their browser cache in time. Other students reported the effectiveness of instructor supervision in the use of the system and felt that instructor supervision facilitated their completion of practice tasks. Moreover, mainly students with better practical skills but poorer performance in theory exams approved the system's facilitation of skills exams because they felt that OMEDT and the OMEDT system could more fairly evaluate their abilities.

Table 1. Feedback from long-term users of the Objective Manipulative Skill Examination of Dental Technicians system.

Item	Mean (SD)	Skewness
The OMEDT[a] system helps me to learn more efficiently.	4.14 (0.990)	–1.479
The OMEDT system can help me to be more efficient in my future work.	4.10 (1.047)	–1.423
The OMEDT system has improved my theoretical knowledge.	3.79 (1.177)	–0.982
The OMEDT system has improved my practical skills.	4.00 (1.035)	–1.038
I was able to use the OMEDT system to study independently without being asked to do so by the instructor.	4.14 (1.060)	–1.840
I may use the system to review in the future in case of interim work or exams.	4.10 (0.939)	–1.607
The OMEDT system is beneficial for dental technicians.	4.14 (0.990)	–1.479
I am willing to take the practical skills exam using the OMEDT system.	4.10 (1.047)	–1.423

[a]OMEDT: Objective Manipulative Skill Examination of Dental Technicians.

XSL•FO
RenderX

Discussion

Principal Findings

Dental technology is characterized as an interface between disciplinary knowledge and the field of professional practice. The teaching, learning, and assessment of dental technology should facilitate connecting discipline-specific theory to laboratory-based practice to the design and fabrication of prostheses. The professional skills necessary to perform such work must be rigorously assessed and periodically examined after certification is obtained to ensure that the practitioner's performance level is always able to meet the needs of the industry. This is the essence of the CDT practical examination. Unfortunately, conducting the CDT practical examination is a global problem. The United States does not have a national certification system for dental technicians. The title of CDT is necessary for educating prospective CDTs and to manage a certified dental laboratory. This is obtained by passing the CDT examination, including written and practical examinations, offered by the National Board for Certification in Dental Laboratory Technology established by the National Association of Dental Laboratories [46]. The United Kingdom has certifications for dental technicians and clinical dental technicians but does not have the CDT examination [47]. New Zealand has a certification, including a written examination for dental technicians, which is obtained by passing the New Zealand Dental Technology Registration Examination offered by the Dental Council [48]. In China, to qualify as a CDT, one must pass the CDT examination provided by the Ministry of Health and Welfare [40]. As in the United Kingdom, this is a written examination only. It is challenging to organize a real material-based examination for test purposes. For example, the whole casting process of metal crown fabrication usually needs more than 12 hours. Owing to the long preparation time and expensive material consumption, it is difficult for the CDT practical examination institutions to prepare semifinished casting products for each candidate. The examination results largely depend not only on the professional performance of the candidates but also on the quality of those semifinished casting products. Considering the potential for fraud and ethical issues, it is unacceptable to examine the final product's quality only or send the material package and require candidates to prepare semifinished products themselves for the CDT practical examination (which is the American plan) [46].

Some possible solutions based on virtual simulation technology have been applied in dental education. Liu et al [49,50] reported that the application of VLNP and RDTES demonstrated some promising features for students' training in preclinical operative and ceramic crown preparation. Osnes et al [51] reported applying a haptic device (Simodont, Nissin Dental Products Europe BV) in the prosthodontics and endodontics areas separately. Different studies [52,53] have shown that Simodont significantly improves students' psychomotor skill performance in these areas. Studies [54-56] have shown the application of a haptic device named Virteasy Dental, and the training of psychomotor skills by using Virteasy Dental can be extended to the fields of implantology and endodontics. Considering the association between dental technology and the stomatology specialty, the performance of virtual simulation technology in dental technology education can be optimistically predicted.

The limitation of the existing virtual simulation products is that most of them are aimed at a single discipline in stomatology. Haptic devices and VLNP and RDTES are separate and lack integration. Moreover, there are no well-known educational or examination standards for dental simulators or associated exercises [20,39,57]. When virtual simulation technology was attempted for the CDT practical examination, existing products were unfortunately not targeted. The OMEDT system, as the world's first virtual simulation system based on professional requests for dental technology studies, can fill the gap of virtual simulation education in dental technology. Virtual simulation technology limits the length of time in a single examination station, and the V-H-R architecture ensures the diversity in the examination content. The examination mode based on OSCE ensures that students' professional skills are assessed completely. The effectiveness of OMEDT and the OMEDT system was verified by student evaluations. Most of the students gave positive comments on the usefulness of the OMEDT system for their studies and examinations and believed that the OMEDT system helped them evaluate their professional ability more fairly and comprehensively. Further, students expressed a strong desire for self-directed learning and continuing education during work, demonstrating that the OMEDT system can be used as part of the CDT practical examination.

The V-H-R architecture (1 program module) of OMEDT can be used for multiple curricula. For example, D-RPD can be used as a tool by educators to demonstrate and explain the Kennedy Classification during theory lectures on removable denture technology. D-RPD can also be used as a tool by educators to present and explain the design of RPDs or by students to complete the design task of the RPD during their experimental classes on RPD technology. Further, D-RPD can be used as a tool by students during their internship and graduation to practice freely and as a tool for instituting graduation exams. The OMEDT system can be used by students from their first exposure to professional knowledge and continue to be used in the final vocational examination, which could help unify teaching objectives and examination standards. Thus, the OMEDT system is an integrated solution for the CDT practical examination. China may eventually be able to institute a formal CDT practical examination based on the OMEDT system.

Our study also extends the concept and application of RDTES technology. Traditionally, RDTES devices are often described as 2D or 3D image acquisition devices [58] that record and store the working process of students by using a simulator [59]. Whether this technology actually improves dental students' cognitive ability and psychomotor skills is still as matter of debate. In the field of dental technology education, the practical significance of this technology is weakened, because most of the operations involved in the dental laboratory are not independent, but rather need as many as a dozen processes—a total time of approximately 5-7 days is required. This is part of the reason that the dental virtual simulator cannot be directly used in dental technology, as mentioned in the previous discussion. Considering the characteristics of dental technology specialties, it is important to evaluate and observe students in

XSL•FO
RenderX

real time from the knowledge of the results [58] dimension and reconstruct the concept and logic of RDTES technology in the field of dental technology. According to "A scoping review of simulation-based dental education" by Dr Higgins [60], the training of aspiring dentists by using simulation-based technology can be done in phases. These phases include briefing, simulation, feedback, debriefing, reflection, and evaluation. Diversity in students' learning styles and motivation is the crucial challenge that course designers face. The RDTES part of the OMEDT system draws on and develops this teaching theory. In addition to the traditional RDTES design, the application of the OMEDT system in teaching practice provides a new evaluation idea for the process evaluation of skills. In conventional dental education, especially in the cultivation of skills, educators are more inclined to evaluate students' skills based on the results. The limitations and inaccuracies of this evaluation standard are apparent. The OMEDT system provides educators with the perspective of growth curve observation, carries out the process assessment of students' psychomotor skills, and can be predicted to be more helpful for cultivating students' psychomotor skills. Educators can more easily make timely adjustments and provide prompt responses during long large-scale training sessions. This will be a new dimension of evaluation of the knowledge of results. This can be illustrated briefly by the application of D-RPD. In the general pedagogical experience, the knowledge point of D-RPD has been set in the concept that 3 class exercises were enough. However, the growth curve analysis provided by the OMEDT system led to other conclusions. It takes more than 5 rounds of continuous design training to show the training effect, which suggests that educators may need to adjust the setting of courses or change the routine habit of assessments. Meanwhile, the observation perspective based on the time dimension of D-Manager helps educators pay timely attention to students' learning motivation and feedback in the continuous task. For example, in the fifth week, the completion time increased significantly. At the same time, the average score decreased, suggesting that educators should pay attention to and adjust the task's difficulty or find the possible problems in students' learning. The time dimension also references the formulation of OMEDT's section time. For example, the validity evaluation experiment of D-RPD suggests that most students can complete a D-RPD within 30-40 minutes, which is reflected in the architectural design of the OMEDT station (D-RPD station is set at 45 minutes). The introduction of the OMEDT system in the dental technology curriculum and the utilization of its data to stratify dental technology students and predict their psychomotor skill performance provide the opportunity to tailor the learning process to meet the individual diversity in students' expertise and allow students to work at their own pace. In this context, the major core curriculum of dental technology could provide an education that leads to the optimal performance of each student. Computer engineers play an indispensable role in the development of new systems. In the process of the experiments, the engineers helped to maintain the system during running time and to collect training data. In addition, the engineers could help students develop good habits of virtual simulation technology application (such as timely cleaning of browser cache), so that students can use the system to study and test more smoothly.

Limitations and Future Work

The main limitation of this study is that owing to time limitations, a comprehensive effectiveness evaluation was not performed for all modules of the OMEDT system. This limitation will be addressed in subsequent studies. In addition, some modules of the OMEDT system were found to have room for improvement in teaching practice. These modules will be improved in the subsequent research plan. OMEDT and the OMEDT system mainly examine the cognitive ability and psychomotor skills in professional performance but not the emotional skills of the dental technicians. The empathy and communication ability of dentists and dental technicians are important factors that determine the success rate of oral clinical treatment. OMEDT and the OMEDT system should continue to develop and explore the scope of this framework in the field of dental interprofessional education. In the future, the development of a professional performance training system for students with stomatology and dental technology majors should be further discussed. The practical application of interprofessional education in dental health education will become the research frontier in the future.

Conclusion

As the world's first virtual simulation system for dental technology education, the OMEDT system integrates 3 existing virtual simulation technologies, that is, VLNP, a haptic interaction, and RDTES. In a virtual environment that saves cost, resources, and manpower, the OMEDT system provides students with a "virtual laboratory" for repeated independent practice. Large-scale data collection provides and validates new dimensions for educators to observe students' professional performance in experimental teaching activities, which is of positive significance to the assessment of students' learning process. The OMEDT system has positive potential as an overall solution to the CDT practical examination. In the future, we will improve OMEDT according to the collected data and conduct more experiments to evaluate the training effectiveness of other modules in the OMEDT system.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 31871464 for PJ) and the Teaching Reformation Fund of the College of Stomatology, Chongqing Medical University (KQJ201810 and KQJ202003 for MP). All authors thank Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd for assisting in the development of the Objective Manipulative Skill Examination of Dental Technicians system.

XSL•FO
RenderX

Authors' Contributions

MP designed the structure of the Objective Manipulative Skill Examination of Dental Technicians (OMEDT) and the OMEDT system, designed the validity experiment, collected and analyzed the data, and drafted the paper. XZ helped complete the software algorithm part of the OMEDT system and revised a part of the manuscript. YD and DL helped collect and analyze the data. LJ assisted in organizing the validity experiment. JL assisted in organizing the structure of the manuscript, conceived the data, and revised the manuscript. PJ assisted in revising the manuscript. All authors approved the manuscript.

Conflicts of Interest

None declared.

References

1. Joda T, Yeung AWK, Hung K, Zitzmann NU, Bornstein MM. Disruptive Innovation in Dentistry: What It Is and What Could Be Next. J Dent Res 2021 May;100(5):448-453. [doi: 10.1177/0022034520978774] [Medline: 33322997]
2. Zheng L, Yue L, Zhou M, Yu H. Dental laboratory technology education in China: current situation and challenges. J Dent Educ 2013 Mar;77(3):345-347. [doi: 10.1002/j.0022-0337.2013.77.3.tb05476.x] [Medline: 23486899]
3. Evans J, Dirks S. Relationships of Admissions Data and Measurements of Psychological Constructs with Psychomotor Performance of Dental Technology Students. Journal of Dental Education 2001 Sep 01;65(9):874-882 [FREE Full text] [doi: 10.1002/j.0022-0337.2001.65.9.tb03435.x] [Medline: 11569603]
4. Newell KM. Motor skill acquisition. Annu Rev Psychol 1991;42:213-237. [doi: 10.1146/annurev.ps.42.020191.001241] [Medline: 2018394]
5. Boot WR, Sumner A, Towne TJ, Rodriguez P, Anders Ericsson K. Applying Aspects of the Expert Performance Approach to Better Understand the Structure of Skill and Mechanisms of Skill Acquisition in Video Games. Top Cogn Sci 2017 Apr;9(2):413-436 [FREE Full text] [doi: 10.1111/tops.12230] [Medline: 27868367]
6. Wajngarten D, Pazos JM, Menegazzo VP, Novo JPD, Garcia PPNS. Magnification effect on fine motor skills of dental students. PLoS One 2021;16(11):e0259768 [FREE Full text] [doi: 10.1371/journal.pone.0259768] [Medline: 34748614]
7. Zitzmann NU, Matthisson L, Ohla H, Joda T. Digital Undergraduate Education in Dentistry: A Systematic Review. IJERPH 2020 May 07;17(9):3269. [doi: 10.3390/ijerph17093269] [Medline: 32392877]
8. Neumann LM. Trends in dental and allied dental education. J Am Dent Assoc 2004 Sep;135(9):1253-1259. [doi: 10.14219/jada.archive.2004.0400] [Medline: 15493389]
9. Bobich AM, Mitchell BL. Transforming Dental Technology Education: Skills, Knowledge, and Curricular Reform. J Dent Educ 2017 Sep;81(9):eS59-eS64. [doi: 10.21815/JDE.017.035] [Medline: 28864805]
10. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg 1995 Dec;222(6):735-742. [doi: 10.1097/00000658-199512000-00007] [Medline: 8526580]
11. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. Medical Teacher 2013 Aug 22;35(9):e1437-e1446. [doi: 10.3109/0142159x.2013.818634] [Medline: 23968323]
12. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. Medical Teacher 2013 Aug 22;35(9):e1447-e1463. [doi: 10.3109/0142159X.2013.818635] [Medline: 23968324]
13. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975 Feb 22;1(5955):447-451 [FREE Full text] [doi: 10.1136/bmj.1.5955.447] [Medline: 1115966]
14. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
15. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health Education Serious Games Targeting Health Care Providers, Patients, and Public Health Users: Scoping Review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
16. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]
17. Avila R, Sobierajski L. A haptic interaction method for volume visualization. 1996 Presented at: Proceedings of Seventh Annual IEEE Visualization '96; October 27-November 1; San Francisco, CA, USA. [doi: 10.1109/visual.1996.568108]
18. Liu L, Zhou R, Yuan S, Sun Z, Lu X, Li J, et al. Simulation training for ceramic crown preparation in the dental setting using a virtual educational system. Eur J Dent Educ 2020 May;24(2):199-206. [doi: 10.1111/eje.12485] [Medline: 31766075]
19. Tang L, Cao Y, Liu Z, Qian K, Liu Y, Liu Y, et al. Improving the quality of preclinical simulation training for dental students using a new digital real-time evaluation system. Eur J Dent Educ 2021 Feb;25(1):100-107. [doi: 10.1111/eje.12580] [Medline: 32777156]
20. Moussa R, Alghazaly A, Althagafi N, Eshky R, Borzangy S. Effectiveness of Virtual Reality and Interactive Simulators on Dental Education Outcomes: Systematic Review. Eur J Dent 2022 Feb;16(1):14-31 [FREE Full text] [doi: 10.1055/s-0041-1731837] [Medline: 34428851]

XSL•FO
RenderX

21. Ziv Stephen D Small Paul Root Wolpe A. Patient safety and simulation-based medical education. Med Teach 2000;22(5):489-495. [doi: 10.1080/01421590050110777] [Medline: 21271963]
22. Lyroudia K, Mikrogeorgis G, Bakaloudi P, Kechagias E, Nikolaidis N, Pitas I. Virtual endodontics: three-dimensional tooth volume representations and their pulp cavity access. J Endod 2002 Aug;28(8):599-602. [doi: 10.1097/00004770-200208000-00009] [Medline: 12184423]
23. Rosenberg H, Grad HA, Matear DW. The Effectiveness of Computer-Aided, Self-Instructional Programs in Dental Education: A Systematic Review of the Literature. Journal of Dental Education 2003 May 01;67(5):524-532. [doi: 10.1002/j.0022-0337.2003.67.5.tb03654.x]
24. Jasinevicius TR, Landers M, Nelson S, Urbankova A. An Evaluation of Two Dental Simulation Systems: Virtual Reality versus Contemporary Non-Computer-Assisted. Journal of Dental Education 2004 Nov 01;68(11):1151-1162. [doi: 10.1002/j.0022-0337.2004.68.11.tb03860.x]
25. LeBlanc VR, Urbankova A, Hadavi F, Lichtenthal RM. A Preliminary Study in Using Virtual Reality to Train Dental Students. Journal of Dental Education 2004 Mar 01;68(3):378-383. [doi: 10.1002/j.0022-0337.2004.68.3.tb03754.x]
26. Steinberg AD, Bashook PG, Drummond J, Ashrafi S, Zefran M. Assessment of Faculty Perception of Content Validity of PerioSim©, a Haptic-3D Virtual Reality Dental Training Simulator. Journal of Dental Education 2007 Dec 01;71(12):1574-1582. [doi: 10.1002/j.0022-0337.2007.71.12.tb04434.x]
27. Liu G, Zhang Y, Townsend WT. Force modeling for tooth preparation in a dental training system. Virtual Reality 2008 Jun 26;12(3):125-136. [doi: 10.1007/s10055-008-0096-8]
28. Noborio H, Sasaki D, Kawamoto Y, Tatsumi T, Sohmura T. Mixed reality software for dental simulation system. 2008 Presented at: 2008 IEEE International Workshop on Haptic Audiovisual Environments and Games; October 18-19; Ottawa, ON, Canada. [doi: 10.1109/have.2008.4685292]
29. Luciano C, Banerjee P, DeFanti T. Haptics-based virtual reality periodontal training simulator. Virtual Reality 2009 Feb 13;13(2):69-85. [doi: 10.1007/s10055-009-0112-7]
30. Rhienmora P, Haddawy P, Suebnukarn S, Dailey M. Providing objective feedback on skill assessment in a dental surgical training simulator. 2009 Presented at: Conference on Artificial Intelligence in Medicine in Europe; June 11-14; Berlin, Heidelberg. [doi: 10.1007/978-3-642-02976-9_43]
31. Suebnukarn S, Phatthanasathiankul N, Sombatweroje S, Rhienmora P, Haddawy P. Process and outcome measures of expert/novice performance on a haptic virtual reality system. J Dent 2009 Sep;37(9):658-665. [doi: 10.1016/j.jdent.2009.04.008] [Medline: 19481326]
32. Rhienmora P, Haddawy P, Khanal P, Suebnukarn S, Dailey MN. A Virtual Reality Simulator for Teaching and Evaluating Dental Procedures. Methods Inf Med 2018 Jan 17;49(04):396-405. [doi: 10.3414/me9310]
33. Gal GB, Weiss EI, Gafni N, Ziv A. Preliminary Assessment of Faculty and Student Perception of a Haptic Virtual Reality Simulator for Training Dental Manual Dexterity. Journal of Dental Education 2011 Apr 01;75(4):496-504. [doi: 10.1002/j.0022-0337.2011.75.4.tb05073.x]
34. Yoshida Y, Yamaguchi S, Kawamoto Y, Noborio H, Murakami S, Sohmura T. Development of a multi-layered virtual tooth model for the haptic dental training system. Dent Mater J 2011;30(1):1-6 [FREE Full text] [doi: 10.4012/dmj.2010-082] [Medline: 21282895]
35. Tanzawa T, Futaki K, Tani C, Hasegawa T, Yamamoto M, Miyazaki T. Introduction of a robot patient into dental education. European Journal of Dental Education 2012;16(1):e195-e1e9. [doi: 10.1111/j.1600-0579.2011.00697.x] [Medline: 22251346]
36. Hattori A, Tonami KI, Tsuruta J, Hideshima M, Kimura Y, Nitta H, et al. Effect of the haptic 3D virtual reality dental training simulator on assessment of tooth preparation. J Dent Sci 2022 Jan;17(1):514-520 [FREE Full text] [doi: 10.1016/j.jds.2021.06.022] [Medline: 35028078]
37. Kikuchi H, Ikeda M, Araki K. Evaluation of a Virtual Reality Simulation System for Porcelain Fused to Metal Crown Preparation at Tokyo Medical and Dental University. Journal of Dental Education 2018 Aug 24;77(6):782-792. [doi: 10.1002/j.0022-0337.2013.77.6.tb05530.x]
38. Tanzawa T, Futaki K, Kurabayashi H, Goto K, Yoshihama Y, Hasegawa T, et al. Medical emergency education using a robot patient in a dental setting. Eur J Dent Educ 2013 Feb;17(1):e114-e119. [doi: 10.1111/j.1600-0579.2012.00770.x] [Medline: 23279398]
39. Li Y, Ye H, Ye F, Liu Y, Lv L, Zhang P, et al. The Current Situation and Future Prospects of Simulators in Dental Education. J Med Internet Res 2021 Apr 08;23(4):e23635 [FREE Full text] [doi: 10.2196/23635] [Medline: 33830059]
40. The national vocational skill standard for dental technician. National Health Commision of the People's Republic of China. 2021. URL: http://www.gov.cn/xinwen/2020-07/31/content_5531660.htm [accessed 2022-08-25]
41. Wang D, Zhang Y, Hou J, Wang Y, Lv P, Chen Y, et al. iDental: A Haptic-Based Dental Simulator and Its Preliminary User Evaluation. IEEE Trans. Haptics 2012 May 24;5(4):332-343. [doi: 10.1109/toh.2011.59] [Medline: 26964131]
42. Wang D, Zhang X, Zhang Y, Xiao J. Configuration-Based Optimization for Six Degree-of-Freedom Haptic Rendering for Fine Manipulation. IEEE Trans. Haptics 2013 Apr;6(2):167-180. [doi: 10.1109/toh.2012.63]
43. Mitchell EL, Arora S. How educational theory can inform the training and practice of vascular surgeons. J Vasc Surg 2012 Aug;56(2):530-537 [FREE Full text] [doi: 10.1016/j.jvs.2012.01.065] [Medline: 22840904]

XSL•FO
RenderX

44. Digital dental photography. American Academy of Cosmetic Dentistry. 2021. URL: https://aacd.com/central/digital-dental-photography [accessed 2021-05-01]
45. National virtual simulation experiment teaching curriculum sharing platform. iLab-x.com. URL: http://www.ilab-x.com/details/v4?id=4401&isView=true [accessed 2021-05-01]
46. National Board for Certification in Dental Laboratory Technology. URL: https://nbccert.org/certificants/certified-dental-technician [accessed 2022-08-25]
47. Registration for dentists and dental care professionals. The General Dental Council (GDC) in United Kingdom. URL: https://www.gdc-uk.org/professionals/education/recent-inspections/inspections-dental-tech [accessed 2021-05-01]
48. New Zealand dental technology registration examination. The Dental Council in New Zealand. URL: https://www.dcnz.org.nz/i-want-to-practise-in-new-zealand/dental-technicians/eligibility-for-the-nzdtechrex/ [accessed 2021-05-01]
49. Liu L, Li J, Yuan S, Wang T, Chu F, Lu X, et al. Evaluating the effectiveness of a preclinical practice of tooth preparation using digital training system: A randomised controlled trial. Eur J Dent Educ 2018 Nov;22(4):e679-e686. [doi: 10.1111/eje.12378] [Medline: 29952122]
50. Liu L, Zhou R, Yuan S, Sun Z, Lu X, Li J, et al. Simulation training for ceramic crown preparation in the dental setting using a virtual educational system. Eur J Dent Educ 2020 May;24(2):199-206. [doi: 10.1111/eje.12485] [Medline: 31766075]
51. Osnes C, Duke A, Wu J, Franklin P, Mushtaq F, Keeling A. Investigating the construct validity of a haptic virtual caries simulation for dental education. BMJ Simul Technol Enhanc Learn 2021;7(2):81-85 [FREE Full text] [doi: 10.1136/bmjstel-2019-000549] [Medline: 35520380]
52. Towers A, Dixon J, Field J, Martin R, Martin N. Combining virtual reality and 3D-printed models to simulate patient-specific dental operative procedures-A study exploring student perceptions. Eur J Dent Educ 2022 May;26(2):393-403. [doi: 10.1111/eje.12715] [Medline: 34498360]
53. Hattori A, Tonami K, Tsuruta J, Hideshima M, Kimura Y, Nitta H, et al. Effect of the haptic 3D virtual reality dental training simulator on assessment of tooth preparation. J Dent Sci 2022 Jan;17(1):514-520 [FREE Full text] [doi: 10.1016/j.jds.2021.06.022] [Medline: 35028078]
54. Cormier J, Pasco D, Syllebranque C, Querrec R. VirTeaSy: a haptic simulator for dental education. 2011 Sep 15 Presented at: ICVL 2011: The 6th International Conference on Virtual Learning; October 28-29; Cluj-Napoca, Romania.
55. Vincent M, Joseph D, Amory C, Paoli N, Ambrosini P, Mortier, et al. Contribution of Haptic Simulation to Analogic Training Environment in Restorative Dentistry. J Dent Educ 2020 Mar 16;84(3):367-376. [doi: 10.21815/jde.019.187] [Medline: 32176342]
56. Dixon J, Towers A, Martin N, Field J. Re-defining the virtual reality dental simulator: Demonstrating concurrent validity of clinically relevant assessment and feedback. Eur J Dent Educ 2021 Feb;25(1):108-116. [doi: 10.1111/eje.12581] [Medline: 32780500]
57. van Riet TCT, Chin Jen Sem KTH, Ho J, Spijker R, Kober J, de Lange J. Robot technology in dentistry, part two of a systematic review: an overview of initiatives. Dent Mater 2021 Aug;37(8):1227-1236 [FREE Full text] [doi: 10.1016/j.dental.2021.06.002] [Medline: 34162501]
58. Kaufman HH, Wiegand RL, Tunick RH. Teaching surgeons to operate—Principles of psychomotor skills training. Acta neurochir 1987 Aug;87(1-2):1-7. [doi: 10.1007/bf02076007]
59. Mays KA, Branch-Mays GL. A Systematic Review of the Use of Self-Assessment in Preclinical and Clinical Dental Education. J Dent Educ 2016 Aug;80(8):902-913. [Medline: 27480701]
60. Higgins D, Hayes M, Taylor J, Wallace J. A scoping review of simulation-based dental education. MedEdPublish 2020 Feb 27;9:36. [doi: 10.15694/mep.2020.000036.1]

Abbreviations

CDT: certified dental technician
DOF: degrees of freedom
D-RPD: removable partial denture design
OMEDT: Objective Manipulative Skill Examination of Dental Technicians
OSCE: Objective Structured Clinical Examination
RDTES: real-time dental training and evaluation system
RPD: removable partial denture
VLNP: virtual learning network platform

XSL•FO
RenderX

Edited by N Zary; submitted 01.01.22; peer-reviewed by R Ciorap, A Garavand; comments to author 14.04.22; revised version received 31.05.22; accepted 21.07.22; published 12.09.22.

Please cite as:
Pang M, Zhao X, Lu D, Dong Y, Jiang L, Li J, Ji P
Preliminary User Evaluation of a New Dental Technology Virtual Simulation System: Development and Validation Study
JMIR Serious Games 2022;10(3):e36079
URL: https://games.jmir.org/2022/3/e36079
doi:10.2196/36079
PMID:36094803

©Mengwei Pang, Xiaohan Zhao, Daiyu Lu, Yihan Dong, Lin Jiang, Jie Li, Ping Ji. Originally published in JMIR Serious Games (https://games.jmir.org), 12.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Migration of an Escape Room–Style Educational Game to an Online Environment: Design Thinking Methodology

Maja Videnovik[1*], MSc; Tone Vold[2*], MSc; Georgina Dimova[1*], BSc; Linda Vibeke Kiønig[2*], MSc; Vladimir Trajkovik[3*], PhD

[1]Center for Innovations and Digital Education Dig-Ed, Skopje, North Macedonia

[2]Department of Business Administration and Organizational Studies, Inland Norway University of Applied Sciences, Elverum, Norway

[3]Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Skopje, North Macedonia

[*]all authors contributed equally

Corresponding Author:
Tone Vold, MSc
Department of Business Administration and Organizational Studies
Inland Norway University of Applied Sciences
Postbox 400
Elverum, 2418
Norway
Phone: 47 97064109
Email: tone.vold@inn.no

Abstract

Background: The COVID-19 pandemic outbreak has led to a sudden change in education, closing schools and shifting to online teaching, which has become an enormous challenge for teachers and students. Implementing adequate online pedagogical approaches and integrating different digital tools in the teaching process have become a priority in educational systems. Finding a way to keep students' interest and persistence in learning is an important issue that online education is facing. One possible way to establish engaging and interactive learning environments, using the energy and enthusiasm of students for educational purposes, is the use of game-based learning activities and gamification of different parts of the educational process.

Objective: This paper presents a use case of migrating an escape room–style educational game to an online environment by using the design thinking methodology. We wanted to show that the design thinking methodology is useful to create engaging and motivating online games that provide educational value.

Methods: Starting from students' perspective, we created a simple digital escape room–style game where students got an opportunity to self-assess their knowledge in computer science at their own pace. Students tested this prototype game, and their opinions about the game were collected through an online survey. The test's goal was to evaluate the students' perceptions about the implemented digital escape room–style educational game and gather information about whether it could achieve students' engagement in learning computer science during online teaching.

Results: In total, 117 students from sixth and seventh grades completed the survey regarding the achieved student engagement. Despite the differences in students' answers about game complexity and puzzle difficulty, most students liked the activity (mean 4.75, SD 0.67, on a scale from 1 to 5). They enjoyed the game, and they would like to participate in this kind of activity again (mean 4.74, SD 0.68). All (n=117, 100%) students found the digital escape room–style educational game interesting for playing and learning.

Conclusions: The results confirmed that digital escape room–style games could be used as an educational tool to engage students in the learning process and achieve learning outcomes. Furthermore, the design thinking methodology proved to be a useful tool in the process of adding novel educational value to the digital escape room–style game.

(JMIR Serious Games 2022;10(3):e32095) doi:10.2196/32095

KEYWORDS

digital games; escape room; computational thinking; gamification; students' engagement; interactive learning environments; serious games; digital health; education; student education; learning outcomes; digital learning; digital education; educational games

XSL•FO
RenderX

Introduction

Education and teaching nowadays are shifting to a learner-centered approach and skill development where a teacher guides students in the learning process, enabling them to progress at their own pace, considering different students' learning styles. Different digital tools and learning paradigms are often included in the education in order to increase the quality of achieved learning [1].

The outbreak of COVID-19 has led to a sudden change in education. More than a billion students worldwide have been affected by school and university closures due to the COVID-19 pandemic [2,3]. The affected number of students equals around 90% of the world's enrolled students [2,4].

This situation has led shifting to online teaching, which has become an enormous challenge for teachers and students. There has been a remarkable rise in e-learning, whereby teaching is undertaken remotely and on different digital platforms [5]. Teachers have been facing a big challenge in how to adapt teaching materials and methodologies to an online environment. Achievement of educational goals, planning online activities, keeping students' interest and motivation, and assessing gathered knowledge and skills are just some of the issues that teachers have been dealing with [6].

The transitioning period from regular to remote and online teaching created an urgency for teachers' proficiency in digitalization [7]. Students are skilled in taking part in digital lessons, but teachers' development in that sense turns out to be a lot more difficult [8]. Teachers had to become aware of synchronous and asynchronous learning, their characteristics, the benefits and, most importantly, the necessity to combine both to have quality online education.

To achieve necessary flexibility in teaching and learning in an online environment, detailed planning of all activities that should be implemented is needed. Particular focus must be on selecting the appropriate strategies and digital tools that can be used, considering the limited time for a virtual classroom activity. To have more time for tasks acquiring higher-order thinking skills (problem solving, case analysis, or project development) during virtual classes, teachers should plan different asynchronous activities using the flipped classroom approach [9].

A flipped classroom is a pedagogical approach where direct instruction is transferred from the collective learning space to the interactive individual learning space [10]. It proposes to carry out certain learning processes outside the classroom and use the virtual classroom to enhance the learning process by applying the knowledge and concepts creatively [11]. The potential of this approach lies in the fact that the time invested in explaining the subject through lectures can be designated as work that the students can do at home using video recordings, a narrated presentation, text writings, or self-paced games. Thus, teachers can spend more time promoting interaction in the classroom, guaranteeing higher understanding, and reinforcing the content. Asynchronous activities after the virtual class should be designed to reflect on the learning and to empower students' knowledge and skills. They can refer to different activities, such as additional research, solving nonroutine problems, and students' self-assessment. This approach, gaining in popularity, has seen its value grow in the context of forced social distancing, like the situation we are facing during the pandemic.

Finding a way to keep students' interest and persistence in learning is an important issue that online education is facing, especially during asynchronous activities. It is obvious that this educational change must be tied to the tools and resources students use in their everyday lives. Proper integration of such tools and resources during the learning can increase students' interest and motivation in the educational process.

Students' engagement plays a significant role in their achievement within the student-centered learning process [12]. With governments interested in measuring students' outcomes [13] and findings that student engagement can act as a proxy for quality [14], a clear understanding of student engagement becomes essential. Student engagement is a complex process that brings together diverse research threads contributing to student success explanations [15]. It can lead to the acquisition of long-lasting knowledge and skills necessary for living in this constantly changing world.

One possible way for establishing engaging and interactive learning environments, using students' energy and enthusiasm while playing games for educational purpose, is the use of game-based learning activities and gamification of different aspects of the educational process.

Educational games increase students' engagement in the learning process. Digital educational games can create new types of interactive learning experiences by combining both game design and instructional design approaches. These interactive learning experiences can increase students' engagement and speed up the process of achieving learning outcomes in the digital learning environment, similar to the way this is achieved in the face-to-face learning environment [16].

According to Sung and Hwang [17], digital games enable problem-solving, decision-making, and strategic planning skills. By using games, new and powerful learning methods in the classroom can be created [18]. By using games, students explore, question, discover, create, and evaluate essential skills that should be acquired. Analysis of engagement in games can provide valuable insights into game mechanisms that can then be applied to games for learning [19].

Gamification, achieved by adding game elements, such as reward and competition during the learning process, provides visible incentives for students' behavior and can increase cognitive load and achievement levels [20]. It can support learning by providing collaboration and self-guided study and increasing motivation, engagement, creativity, and retention [21]. Although gamification and game-based learning are often considered to be synonyms, these 2 concepts are distinct. Game-based learning refers to adopting a game as an educational tool for learning a specific subject in this context. In contrast, gamification is an educational strategy based on game elements in the learning process [22]. Technology-enhanced game-based learning and gamification can increase positive educational results, especially in challenging learning environments [23-25].

XSL•FO
RenderX

The implementation of learning scenarios that use technological resources in and out of distance learning environments face different challenges. It is challenging to match successful games to the curriculum and to use them in the educational process [26]. To achieve valuable integration of games in education, teachers need to connect pedagogical approaches and entertainment [1]. Using interesting games, with features popular among students, for the development of a fun, educational game that leads to the achievement of learning outcomes is 1 possible approach. For example, a mobile-based treasure hunt educational game using augmented reality [27,28] can improve learning experiences by providing active discovery of facts and balancing physical and digital interactions [29]. To encourage the use of games in learning, it is essential to develop an understanding of the tasks, activities, skills, and operations that the games can offer and examine how these match desired learning outcomes while providing a positive learning experience [30]. Most of the games' usage in the educational process is related to knowledge acquisition and content understanding [31].

An escape room is 1 of the games that are becoming extremely popular among students in the past decade. It is a game where a team of players discovers clues, solves puzzles, and accomplishes tasks in 1 or more rooms to progress and accomplish a specific goal in a limited amount of time [32]. To win the game ("escape"), the players must solve the challenges (puzzles) using hints and clues and developing a strategy [11]. To make this kind of game educational, questions concerning course material can be incorporated within the puzzles. Students then will have to master the material to succeed in the game, which will enhance their learning and increase their interest and engagement in learning [33]. In this way, educational escape room games can improve the skills of students [34]. Escape rooms test the players' problem-solving, lateral thinking, and teamwork skills by providing various puzzles and challenges, making this game perfect for implementation in an educational context [35]. Incorporating educational elements within the challenges leads to achievement of educational goals and increase in students' motivation and interest in learning [36,37]. Educational escape room games are a new type of learning activity with the promise of enhancing students' learning through highly engaging experiences [25]. Escape rooms have drawn educators' attention due to their ability to foster teamwork and creative thinking in an engaging way for students [33,38]. By adding learning activities to the story in a meaningful way, these escape room games can create memorable learning experiences that cannot be replicated by typical classroom activities [35,39].

Educators have introduced so-called escape room games into their teaching or training practices to solve educational content puzzles [40]. An escape room game contextualizes educational content into a meaningful and inspiring experience based on game-based and collaborative learning [39]. This approach has been used for various subjects [41-43] and various educational process levels [25,44,45].

Developing an escape room in a digital environment suitable for online learning [46,47] is the latest trend in this context. This faces additional challenges because it has to be set up in an online environment, but considering that there are no physical limitations, it can involve a large number of students [48]. The effects of educational digital escape rooms on students' learning performance, learning motivation, and problem-solving ability seem positive [34,45,49].

This paper discusses how we used a design thinking methodology approach for providing a new educational value to a digital escape room game. We started from students' attitudes toward an already created and played physical escape room game [12], which enabled collaborative learning in the classroom. The idea was to migrate this escape room game to an online environment in order to keep students' interest and persistence during asynchronous learning. However, due to the new circumstances dictated by the characteristics of an online environment and the short time for its creation, we had to find a new type of engaging experience during the learning process that was not based on collaboration. The digital escape room created was tested by students, and this paper elaborates on whether digital escape room games can increase students' interest and motivation in achieving learning outcomes during remote teaching. Starting from this, the research questions were:

- Can a digital escape room–style educational game be used to achieve student engagement in remote teaching and online learning?
- Can we use the design thinking methodology to create an online escape room–style educational game that is interesting and motivational for students?
- Can we use the design thinking methodology to provide novel educational value to the escape room–style educational game?

The first research question was answered by analyzing students' answers after using a digital escape room–style educational game during the learning process. The second and third questions concerned the used methodology (design thinking) for developing this kind of a game, so they were answered through the results (outcomes) of each phase of the design thinking process.

The next section explains how we used the design thinking methodology to migrate an escape room–style educational game from a physical to an online environment and the methodology we used to evaluate the achieved level of student engagement. In the Results section, the results obtained from the students' surveys focusing on student engagement and achieved educational value are elaborated. A discussion about these results and the limitations of this research, as well as ideas for future work, are presented in the Discussion section. The last section concludes the paper.

Methods

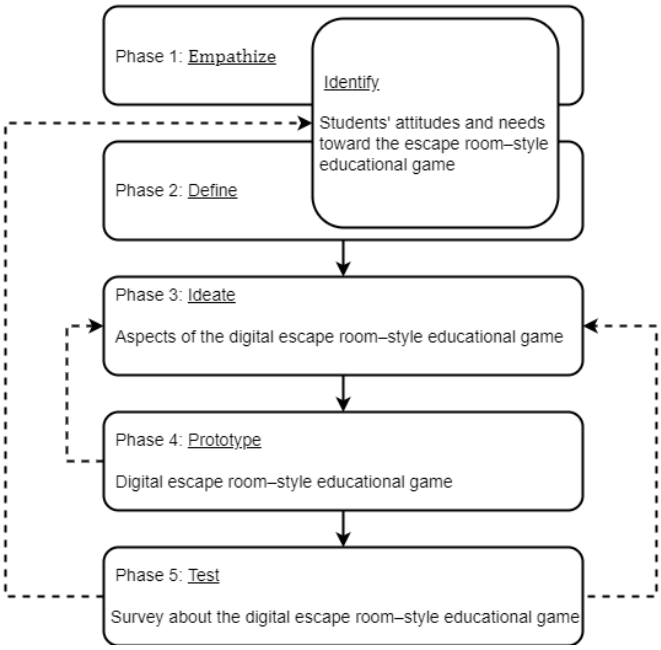
Study Design

We used the design thinking methodology to migrate and transform an escape room–style educational game to an online environment. At the same time, we wanted to check whether a digital escape room game can provide educational value and still be engaging for students, even without teacher supervision and a lack of student collaboration.

XSL•FO
RenderX

The background of design thinking is based on a problem-solving approach [50]—in our case the need to design an online game that will be interesting and will include educational elements at the same time. To achieve this, the digital escape room–style educational game development process was conducted using a modified version of the 5-stage design thinking method proposed by the Hasso-Plattner Institute of Design at Stanford [51-53]. According to this methodological model, the 5 phases of design thinking are *empathize*, *define* (the problem), *ideate*, *prototype*, and *test*. The design thinking methodology applied to our particular case is presented in Figure 1.

Figure 1. Design thinking methodology used in our study.



Design thinking is a popular methodology that inspires a human-centered approach toward design. For our research, we consolidated the *empathize* and *define* phases into 1 phase, the *identify* phase, which gave us an understanding of the problem from the students' perspective and helped us define the problem in a student-centered manner, starting from their needs and attitudes toward the escape room–style educational game.

Design thinking is a cognitive process from which design concepts emerge [54], and its use suggests that we put emphasis on thinking of design as a process. So, in the next stage, *ideate*, we determined aspects of the digital escape room game, starting from students' attitudes and the educational elements that should be included.

In the next, *prototype*, phase, we tried to develop a digital escape room–style educational game and test it with students in the testing phase. Students' feedback from this testing phase will show whether the game corresponds to their needs and whether it is in accordance with the research aims. This feedback will also provide insight into the process and its limitations and can be used for identification of students' needs as a starting point for another design thinking process in order to improve the quality of the final product—the digital escape room–style educational game.

Although we talk about the design thinking process in terms of sequential steps, it is actually a highly iterative loop. With each phase, we can make new discoveries, which may require revision of the previous stages. The solid lines in Figure 1 represent the chronology of the methodology steps, while the dotted lines represent how different outputs of the design thinking methodology can be used [51].

Each of the phases of the design thinking methodology is explained next.

Identify Phase: Identification of Students' Needs and Attitudes Toward the Digital Escape Room–Style Room Educational Game

The first phase of design thinking is to gain an emphatic understanding of the problem we want to solve—how to create a digital escape room–style educational game that will be interesting and educational at the same time and that will engage students in the learning process. To achieve this, we started from students' attitudes toward escape room–style educational games, and after that, we identified the game aspects that should be adapted for an online environment.

To identify students' needs and their attitudes toward playing an escape room–style educational game and to test whether this kind of game-based learning can be used for achieving students' engagement during computer science classes, we used already obtained information from the analyses of students' experiences during the implemented escape room–style educational game in physical surroundings, as described previously [12]. That escape room–style educational game combined computer-based and unplugged puzzles in a computer lab space in order to create a highly engaging activity without compromising its educational value. The puzzles were connected with the course material in general and encouraged research, communication, collaboration, and teamwork. These first findings, presented in detail in our

XSL•FO
RenderX

previous work [12], provided evidence that escape room–style educational games constitute a compelling way to achieve student engagement, so transforming this learning experience to an online environment was not a trivial task. The technological restrictions for nonverbal communication and the lack of free movement in a certain physical space created the need to redesign the whole gameplay.

Ideate Phase: Ideate Aspects of the Digital Escape Room–Style Educational Game

The next phase of the design thinking methodology was to determine necessary aspects of the digital escape room–style educational game. Based on the results of our previous study, we concluded that the collaboration and teamwork that were present in a physical surrounding needed to be replaced with a different aspect of a game that would infuse thrill in a digital escape room game. At the same time, the defined challenges (tasks) needed to have both game-related and novel education-related purposes. To ideate a solution to these nontrivial issues, we started with the results of a survey for applying escape room games in the educational process [12]. Based on those results, we concluded that although the digital version of such game will have no cooperative element, if we manage to keep the thrill and competitiveness, students will still enjoy the game, thus creating a positive and engaging educational experience. To achieve this, we needed to define a new educational value for the game that would replace the collaboration and teamwork required while grasping the concepts of computer science. At the same time, we needed to identify a part of the learning process suitable for individual tasks that could be made more appealing as a game.

Prototype Phase: Prototyping the Digital Escape Room–Style Educational Game

The prototype phase started by finding a platform where the game would be created and played by students. The platform should enable creating a user-friendly, easy-to-use online game so that students do not have any problems while playing it from a technical aspect. For that purpose, we used a platform that was already used successfully for developing digital escape room games in an educational context [34]. The developed prototype of the digital escape room game was given as an asynchronous activity during online learning. Students got an opportunity to play the game at a time that was most appropriate for them. They would analyze the questions in the puzzles, learn from their mistakes, and reach the end of the game at their own pace.

Test Phase: Testing the Digital Escape Room–Style Educational Game

Information about students' opinions concerning the digital escape room–style educational game was gathered using an online survey conducted with the students at the end of the activity. The survey was adapted from a similar one on using an escape room–style educational game for teaching programming in a higher education setting [25], and some elements were added from a survey on the implementation of an escape room–style educational game in a physical classroom [12].

The first part of the survey collected demographic information about the participants (gender, age, and school). The second part was designed to measure students' attitudes toward the game, using a 5-point Likert scale, with answer choices ranging from 1 for strongly disagree to 5 for strongly agree. Information about students' perception about the design of the game, the opportunity to deepen their knowledge, their feelings during play, and their attitude toward this kind of learning activity was obtained. Students got an opportunity to reflect on the game and answer what they like the most, what they did not like, and what can be changed to improve the game through open-ended questions at the end of the survey. In this way, we established a possibility for students to provide inputs for further development of the digital escape room–style educational game (dotted line from Figure 1) by providing constructive feedback about their learning experience.

Ethical Considerations

The surveys conducted with students are part of everyday practices in schools in order to measure students' attitudes toward some new teaching approach, and they are completed anonymously by students.

Results

Main Goal

The main goal of this work was to evaluate students' perceptions of the implemented digital escape room–style educational game and to determine whether it can be used to achieve student engagement in learning computer science. This was done as part of the process of migration of a traditional collaborative puzzle-based escape room game played in the computer lab to an online environment using the design thinking methodology. In the following sections, the outcomes of each phase of the design thinking process are elaborated. The results in the first, *identify* phase, were obtained from previous research [12] concerning using a physical escape room in education, and it was our starting point for this new research topic, as described in the following phases of the design thinking process.

Identify Phase: Identification of Students' Needs and Attitudes Toward the Escape Room–Style Educational Game

The first stage of design thinking was to understand the problem we wanted to solve—how to create a digital game that would be interesting and educational from the students' point of view. We started by identifying popular escape room games among students as part of our previous work. By using a survey, students' attitudes toward using traditional escape room games in education and the elements of the developed escape room–style educational game that students liked the most were identified [12]. The survey results showed that students had a positive overall opinion about the escape room–style educational game and thought it was a fun experience. It was a good starting point for planning the future use of escape room–style games in the educational context, as we wanted to start from students' attitudes toward the game.

XSL•FO
RenderX

Most of the students were competitive during participation in the game, and they were impatient to open the next puzzle, which increased their interest, motivation, and persistence during the game. This competitiveness led to students' active engagement during the game, impatiently trying to be the first to finish the game. These findings provide evidence that escape room–style educational games constitute a compelling way to increase student engagement.

Regarding learning effectiveness, students stated that the escape room game helped them improve their knowledge of computer science and that they prefer the escape room game over other educational methods in computer science classes and think that they can learn more with the game than during regular classes. These were essential findings showing that educational elements can be added in escape room–style games, which can lead to the achievement of needed educational outcomes in a fun and exciting way.

These results were consistent with previous studies, which found that escape room–style educational games can improve students' knowledge of a specific topic [33,37]. However, research shows that escape room–style educational games have been little used, in both compulsory and higher education [55], and that there is not much research concerning the use of the escape room–style educational games in computer science in primary education.

Open-ended questions led to a conclusion that students had enormous interest and motivation in participating in the escape room–style educational game, and most students gave positive comments that they thoroughly enjoyed the game experience and would like more similar future activities. These findings and our previous work [12] gave us a solid base for the next phase of design thinking (*ideate*) to determine aspects of the digital escape room–style educational game.

Ideate Phase: Ideating Aspects of the Digital Escape Room–Style Educational Game

After understanding students' attitudes and analyzing and synthesizing the feedback provided by students, which enabled approaching a problem from the students' perspective, the next phase in the design thinking process, an idea for creating a digital online game, emerged.

Starting from students' positive attitudes toward using escape room games in education and students' competitiveness while playing the games, we created a digital escape room–style game that would raise students' individual critical-thinking and problem-solving skills, enabling their active participation in the activities.

We concluded that self-assessment is a great candidate for such a process. Self-assessment is an individual learning process, which is important from an educational point of view since it points out the parts of materials that needs more attention by students. In contrast, to serve its purpose, students need to be engaged in the self-assessment process.

The chosen topic to lay behind the story of the digital escape room–style game was connected to algorithmic thinking because the topic is difficult to understand in a traditional teaching environment and should be learned in a fun and interactive way, where students can progress at their own pace according to their previous knowledge. So, learning outcomes connected to algorithmic and computational thinking were planned to be assessed by this digital escape room–style educational game. The educational elements that were supposed to be implemented mostly acquire higher-order thinking skills (creativity, algorithmic thinking, cooperation, critical thinking, and problem solving) [56]. Implementation of these outcomes in coding gives students an opportunity chance to solve problems and to experiment creatively [57]. Students understand that they can learn from their mistakes, which increases their confidence in a fun and exciting way. Developing these skills makes the topic difficult to teach and learn in a traditional classroom and requires more time for learning and enabling students to progress at their own pace. According to this, creating a simple digital escape room–style game where students can solve different questions connected to algorithmic thinking and get an opportunity to self-assess their knowledge and learn from their mistakes was the next step that was carried out in the proposed methodology.

Prototype Phase: Prototyping the Digital Escape Room–Style Educational Game

A platform that enables creating interactive animated content, Genially [58], was used to create a digital escape room–style educational game. Genially is a platform for creating visually appealing, engaging, interactive digital contents, games, breakout rooms, quizzes, and portfolios. Students could access the game just by following the link given by the teacher, and they could immediately start playing the game. The start page of the created digital escape room–style game is shown in Figure 2.

Students read the story behind the escape room and start the game. The overall theme of the created escape room is finding the secret to obtaining a higher grade in computer science on a given topic, which is hidden in a coffin. The coffin can be opened using a 5-digit number; each number represents a solution to a different puzzle. The sequence of numbers is important and determines the order of solving the puzzles, so the escape room consists of 5 puzzles hidden in 5 different places, which should be solved. Students virtually travel the world from place to place (Figure 3) and solve puzzles to discover all 5 digits.

Identification of the educational elements to be incorporated in the game was conducted. Each puzzle is a question connected to a given algorithm (Figure 4). Puzzles were created carefully, increasing their complexity while going through the game. The first puzzle is the simplest, and each subsequent one uses elements from the previous algorithms. Students analyze each question, pass through the algorithm, and obtain the result—a digit. Finally, students get a 5-digit number used to open a coffin, which is “escape” from the room and end of the game.

This activity was in asynchronous form after the virtual classes. Students implemented the gathered knowledge and skills concerning the topic in concrete situations and got an opportunity to self-assess their knowledge. When playing the digital escape room–style educational game, there was no time restriction. Students passed through the activity at their own pace, at the time that they choose. The questions in the game

XSL•FO
RenderX

are time-consuming since they are connected with the development of higher-order thinking skills, which needs time, so there should be no time restriction. However, time was measured, and students knew that they must keep in mind the time necessary to finish the game. The teacher can monitor the ratio of solved tasks in a certain time. This game design was in line with both initial escape room–style educational games and the findings of the educational room student survey (dotted lines in Figure 1).

Figure 2. Start page of a digital escape room–style educational game.

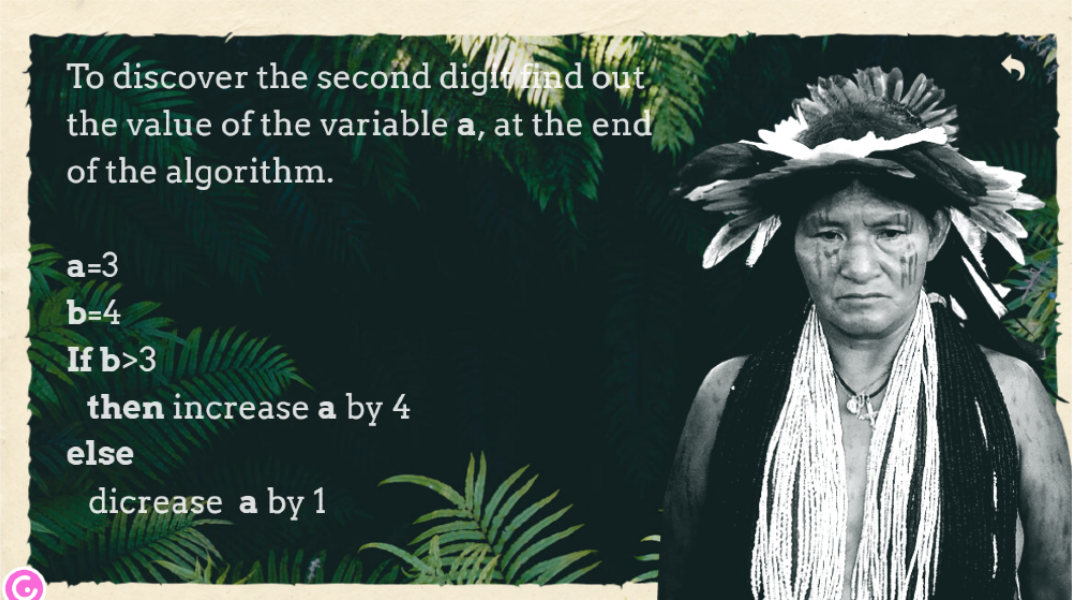


Figure 3. Story behind the puzzles.



XSL•FO
RenderX

Figure 4. Example of a game's puzzle.



Testing Phase: Testing the Digital Escape Room–Style Educational Game

Students answered questions concerning their attitudes about using the digital escape room–style game in an educational context after finishing the game. In total, 117 students from sixth and seventh grades completed the survey to gather information about the achieved student engagement (final phase of the design thinking process), and the results are presented in Table 1. For each question, the mean and SD are presented, showing the dispersion of students' answers. Of the 117 students who played the game, 66 (56.4%) were male and 51 (43.6%) were female. There was no significant difference between the number of students in different classes.

Table 1. Results of students' attitudes concerning using the digital escape room–style game in an educational context.

Question	Mean (SD)
It was easy to follow the instructions of the game.	4.44 (0.86)
The puzzles in the game were not difficult to solve.	4.10 (1.17)
The escape room was interesting for me.	5.00 (0)
I was impatient to open the next puzzle and reach the end.	4.41 (1.12)
I learned something new with the game.	4.26 (1.18)
I liked the activity.	4.74 (0.68)
Learning with this kind of activity is more interesting.	4.74 (0.69)
I would like to participate in this kind of activity again.	4.75 (0.67)
The game was complicated.	1.68 (1.19)
I was encouraged to think by the puzzles.	4.64 (0.76)

The results showed that students did not have any problems while playing the online game; they stated that it was easy to follow the game's instructions (mean 4.44, SD 0.86). However, some students faced problems while solving the puzzles, although most of them stated that the puzzles in the game were not difficult to solve (mean 4.10, SD 1.17).

The number of students that found the game puzzles difficult was bigger in digital than in physical surroundings compared to the results of previous research [12]. This was expected because the aim of the puzzles in the digital escape room was not just to engage students but also to involve them in self-assessment, and students had to demonstrate certain knowledge and skills related to the topic to solve the puzzles. Furthermore, the puzzles in the digital escape room require higher-order thinking skills, not just a repetition of facts, because they are connected to 1 of the most difficult themes for students—algorithmic thinking. This requires computational thinking skills, and that is why they are difficult for some of the students. In contrast, the puzzles in the escape room game in physical surroundings are more focused on establishing collaboration and are easier to solve. The SD values showed differences in students' answers, most probably depending on their acquired knowledge about the topic. The results of game complexity (mean 1.68, SD 1.19) are in line with this finding as well.

XSL•FO
RenderX

Interestingly, students were encouraged to think by the puzzles, and they were motivated to think and engage in solving the puzzles (mean 4.64, SD 0.76). So, we could conclude that although some of the students found the puzzles challenging to solve, they were encouraged to think, by which their higher-order thinking skills were developed. Furthermore, they were impatient to open the next puzzle in the escape room game and reach the end (mean 4.41, SD 1.12). The SD values presented a difference in the students' answers and showed that students had different opinions despite a high average value. Once again, this might be the result of the different levels of students' knowledge. Students who did not have problems solving the puzzles were impatient to finish the game sooner, but for students who did not understand the questions, it was hard, and they spent more time on some of the puzzles. After analyzing students' answers, we found that of 117 students who completed the survey, 90 (76.9%) stated that they did not have problems solving the puzzles and just 16 (13.7%) had some problems in understanding the puzzles' questions.

Despite the differences in students' answers about game complexity and puzzles' difficulty, most students liked the activity (mean 4.74, SD 0.68). They enjoyed the game, and they would like to participate in this kind of activity again (mean 4.75, SD 0.67). All 117 (100%) students found the escape room–style game interesting for playing and learning. Thus, we concluded that implementing a digital escape room–style game during online teaching raises students' interest and motivation.

The students' response concerning the educational element of the digital escape room–style game was that they learned something new (mean 4.26, SD 1.18), once again indicating a difference among the students according to their previous experience in the topic and gathered knowledge and skills. Most students found learning with this kind of activity more attractive than other learning forms (mean 4.74, SD 0.69), confirming that interactive learning activities with gaming elements attract students' attention and keep their engagement during the activity.

Discussion

Principal Findings

The process of migration of the digital escape room-style game presented in this paper elaborates that the lack of face-to-face collaboration used in knowledge acquisition can be replaced by using an escape room game in different parts of the learning process, for example, self-assessment. In this way, we managed to maintain a high level of student engagement during asynchronous activities, the feature of escape room games that seems to be related to an increase in students' skills [34].The initial results from the survey related to the use of the digital escape room–style game, as an asynchronous activity, showed that students found the game easy to play, exciting, and engaging and that they would like similar such activities that will last longer and have more challenging puzzles. A few students had some problems while solving the puzzles, but they also found the activity enjoyable.

Students' answers to the open-ended questions showed that they found the game engaging. Students liked the whole game, moving through the puzzles, logical thinking, and finding the final solution. They mentioned that they learned something new and enjoyed while doing it, which was a completely new experience. They liked that this assessment activity was more like a game, and they had fun while doing it. Learning while playing was stated as 1 of the things that students liked the most. Some of the students liked the escape room's story, how the questions were made and given in the puzzles. They liked the mystery, excitement, and unpredictability and reaching the end of the game. They stated that they liked the overall experience. When students were asked what they liked the least in the online game, they stated that there were a small number of questions and so the game ended quickly. Some of them mentioned that some of the questions were difficult to solve, but they enjoyed the overall experience. These findings are in line with the values of the SD for questions concerning the puzzles' difficulty.

Students said they would like to have more questions of this type for future activities, more puzzles, and even more questions with different complexity levels. Most of the students said they would enjoy more extended challenges. This shows that students enjoyed the game. They would like similar activities soon and even have ideas on how this game can be made more complex and challenging.

All these findings lead to the conclusion that students find our game inspiring and motivational, which confirms that digital escape room–style educational games can be used to achieve student engagement in remote teaching and online learning (our first research question).

Using a design thinking approach, we changed our research focus from being problem oriented to being solution focused and action oriented toward creating a digital escape room–style educational game that would link pedagogical approaches and entertainment elements, starting from students' needs and attitudes toward this topic. During the process, a well-received digital educational game was designed, which was confirmed by students. All students found the game interesting. The analysis of open-ended questions showed that students liked the game, they found it interesting, and they liked solving the puzzles and moving through the game. This is a confirmation that using a design thinking approach, engaging digital educational games that raise students' interest and motivation can be developed (our second research question).

This study elaborates on how design thinking can be used not only to solve a problem (migration of a game from a traditional to an online learning environment), while preserving its most appreciated elements (engagement) but also, by using the same approach, to reach a new educational value, in our case student self-assessment for the learning process (our third research question).

Limitations and Future Work

The main drawback of the migration of an escape room game to an online environment is that due to the COVID 19 pandemic circumstances, students have different types of internet infrastructure at their disposal, so the collaborative value of the escape room game could not be implemented in the short period that was available for the migration of the escape room game.

XSL•FO
RenderX

There is evidence in the literature [45] of successful deployment of collaboration within a digital escape room–style educational game related to software engineering in higher education, but it should be considering that to create the escape room–style educational game, several highly interactive ad hoc web apps had to be developed from scratch by the course teachers. This was confirmed by López-Pernas et al [43] in a comparative study of the results obtained by face-to-face collaboration and a digital escape room–style educational game used to learn programming. There, the students who participated in the remote escape room–style educational game, due to a lack of a common device, requested and obtained nearly twice as many hints as those who participated in the face-to-face collaboration.

Simple transfer of the physical escape room–style educational game to an online environment using Google Forms and videoconferencing tools was presented by Ang et al [59]. Although students enjoyed the digital escape room game, they felt that a real-life escape room would be more fun and interactive.

The possibility for a large class to use an escape room game simultaneously in or outside the usual classroom was presented by Monnot et al [48]. The authors discussed how the escape room game can also be used as a course support or as preparation for a course, such as work at home, but did not present any evidence of the benefits of this idea.

The main limitation of our study is that only 1 instrument, a student survey, was used to evaluate the effects of the digital escape room–style educational game in the learning process. Although student engagement can be determined using surveys, and the literature confirms that the use of educational games can lead to improved student skills [34], further research related to the achievements of educational outcomes can be used to confirm the stated engagement.

Additional limitations are related to the students' age and cultural background. Further research is needed to check whether the same engagement based on digital games can be established with younger age groups or with students from different cultural backgrounds.

Our future research will focus on the ways in which escape room games support collaboration, trust, and reflection, thus fostering a dynamic and flexible learning environment [60]. For this purpose, we will investigate the influence of new technologies in establishing such educational games and education in general, such as ubiquitous computing, pervasive computing, virtuality continuum, ambient intelligence, and wearable computing [61]. We will investigate how using escape room games for student assessment can influence students' educational achievements, considering both negative [62] and positive [63] aspects of this approach.

Conclusion

Student engagement during the learning process is an important element that influences students' learning and determines achieved knowledge and gathered skills. Integration of games that combine entertainment and educational elements during the learning has been recognized as a vital factor for increasing students' interest and motivation during learning. Our previous work confirmed that escape room–style educational games could be used as an engagement activity for students in the classroom environment. Since we have transferred education online due to the pandemic, it was necessary to see whether we can migrate this type of game to an online environment. It was not a trivial process since the main educational benefits of such a game, student engagement through collaboration and teamwork, needed to be transformed and changed into something new in a short period.

In this paper, we presented the process of migration of an escape room–style educational game to an online environment by applying the design thinking methodology. We started with the students' opinion about the use of an escape room–style educational game for learning. Using their opinion, we created a simple digital escape room–style educational game where students got an opportunity to self-assess their knowledge in computer science at their own pace. After playing the game, students' opinions were collected and used to evaluate the process of migration of the escape room–style educational game to an online environment. By using a design thinking approach, we managed to migrate an educational game that links pedagogical approaches and entertainment elements to an online environment and added additional value to it (student self-assessment). The results clearly show that the digital escape room–style educational game was accepted as highly interesting and can be used in remote teaching. The student survey results confirmed that this approach is beneficial for both students and educators. These findings provide a base for our future work. We will explore how digital escape room games can create collaboration and what the influence of new technologies is in establishing this type of educational games. We will focus on escape room games for student assessment and their positive and negative influence on students' educational achievements. We believe that this kind of research can provide a completely new view of pedagogical approaches applicable in distance learning environments.

Acknowledgments

This research was funded by the Inland Norway University of Applied Sciences, Norway. VT acknowledges the Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, for administrative support. MV and GD acknowledge members of the Center for Innovation and Digital Education who encouraged their students to take part in this study.

XSL•FO
RenderX

Conflicts of Interest

None declared

References

1. Videnovik M, Trajkovik V, Kiønig L, Vold T. Increasing quality of learning experience using augmented reality educational games. Multimed Tools Appl 2020 Jun 14;79(33-34):23861-23885 [FREE Full text] [doi: 10.1007/s11042-020-09046-7]
2. UNESCO. COVID-19 Education Response. URL: https://en.unesco.org/covid19/educationresponse/globalcoalition [accessed 2020-12-12]
3. UNICEF. UNICEF and Microsoft Launch Global Learning Platform to Help Address COVID-19 Education Crisis. URL: https://www.unicef.org/press-releases/unicef-and-microsoft-launch-globallearning-platform-help-address-covid-19-education [accessed 2020-12-12]
4. UNESCO. COVID-19 Educational Disruption and Response. URL: https://en.unesco.org/covid19/educationresponse [accessed 2020-12-12]
5. Ferdig R, Baumgartner E, Hartshorne R, Kaplan-Rakowski R, Mouza C. Teaching, Technology, and Teacher Education during the COVID-19 Pandemic: Stories from the Field. Waynesville, NC: Association for the Advancement of Computing in Education; 2020.
6. Chertoff N, Thompson A. Applied Strategies for Remote Student Teaching due to COVID-19. URL: https://academicworks.cuny.edu/oaa_pubs/18/ [accessed 2021-01-06]
7. van der Spoel I, Noroozi O, Schuurink E, van Ginkel S. Teachers' online teaching expectations and experiences during the Covid19-pandemic in the Netherlands. Eur J Teacher Educ 2020 Sep 17;43(4):623-638 [FREE Full text] [doi: 10.1080/02619768.2020.1821185]
8. Aslan A, Zhu C. Pre-service teachers' perceptions of ICT integration in teacher education in Turkey. Turkish Online J Educ Technol 2015;14(3):97-110.
9. Izagirre-Olaizola J, Morandeira-Arca J. Business management teaching-learning processes in times of pandemic: flipped classroom at a distance. Sustainability 2020 Dec 04;12(23):10137 [FREE Full text] [doi: 10.3390/su122310137]
10. Flipped Learning Network. What Is Flipped Learning?. URL: https://flippedlearning.org [accessed 2021-01-06]
11. Kim D, Yao J. A treasure hunt model for inquiry-based learning in the development of a web-based learning support system. J UCS 2010;16(14):1853-1881 [FREE Full text] [doi: 10.1007/978-1-84882-628-1_5]
12. Dimova G, Videnovik M, Trajkovik V. Using educational Escape room to increase students' engagement in learning computer science. 2020 Presented at: CIIT 2020: 17th International Conference on Informatics and Information Technologies; May 8-9, 2020; North Macedonia.
13. Zepke N, Leach L. Beyond hard outcomes: 'soft' outcomes and engagement as student success. Teach Higher Educ 2010 Dec;15(6):661-673 [FREE Full text] [doi: 10.1080/13562517.2010.522084]
14. Kuh G. What Student Affairs Professionals Need to Know About Student Engagement. Journal of College Student Development 2009;50(6):683-706 [FREE Full text] [doi: 10.1353/csd.0.0099]
15. Fredricks J, Blumenfeld P, Paris A. School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research 2016 Jul 01;74(1):59-109 [FREE Full text] [doi: 10.3102/00346543074001059]
16. Neumann K, Alvarado-Albertorio F, Ramírez-Salgado A. Online approaches for implementing a digital escape room with preservice teachers. J Technol Teacher Educ 2020;28(2):415-424 [FREE Full text]
17. Sung H, Hwang G. A collaborative game-based learning approach to improving students' learning performance in science courses. Comput Educ 2013 Apr;63:43-51 [FREE Full text] [doi: 10.1016/j.compedu.2012.11.019]
18. Glover I. Play as you learn: gamification as a technique for motivating learners. 2013 Presented at: EdMedia 2013 - World Conference on Educational Media and Technology; 2013; Victoria, Canada p. 1999-2008.
19. Boyle E, Hainey T, Connolly T, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Comput Educ 2016 Mar;94:178-192 [FREE Full text] [doi: 10.1016/j.compedu.2015.11.003]
20. Turan Z, Avinc Z, Kara K, Goktas Y. Gamification and education: achievements, cognitive loads, and views of students. Int J Emerg Technol Learn 2016 Jul 21;11(07):64 [FREE Full text] [doi: 10.3991/ijet.v11i07.5455]
21. Dichev C, Dicheva D. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. Int J Educ Technol High Educ 2017 Feb 20;14(1):9 [FREE Full text] [doi: 10.1186/s41239-017-0042-5]
22. Trajkovik V, Malinovski T, Vasileva-Stojanovska T, Vasileva M. Traditional games in elementary school: relationships of student's personality traits, motivation and experience with learning outcomes. PLoS One 2018;13(8):e0202172 [FREE Full text] [doi: 10.1371/journal.pone.0202172] [Medline: 30125300]
23. Alzubi T, Fernandez R, Flores J, Duran M, Cotos J. Improving the working memory during early childhood education through the use of an interactive gesture game-based learning approach. IEEE Access 2018;6:53998-54009 [FREE Full text] [doi: 10.1109/access.2018.2870575]
24. Marín B, Frez J, Cruz-Lemus J, Genero M. An empirical investigation on the benefits of gamification in programming courses. ACM Trans Comput Educ 2019 Mar 31;19(1):1-22 [FREE Full text] [doi: 10.1145/3231709]

XSL•FO
RenderX

25. Lopez-Pernas S, Gordillo A, Barra E, Quemada J. Examining the use of an educational escape room for teaching programming in a higher education setting. IEEE Access 2019;7:31723-31737 [FREE Full text] [doi: 10.1109/access.2019.2902976]
26. Videnovik M, Kiønig L, Vold T, Trajkovik V. Testing framework for investigating learning outcome from quiz game: a study from Macedonia and Norway. 2018 Presented at: 17th International Conference on Information Technology Based Higher Education and Training (ITHET); 2018; Olhao, Portugal p. 1-5 URL: https://doi.org/10.1109/ithet.2018.8424777 [doi: 10.1109/ithet.2018.8424777]
27. Barma S, Daniel S, Bacon N, Gingras M, Fortin M. Observation and analysis of a classroom teaching and learning practice based on augmented reality and serious games on mobile platforms. Int J Serious Games 2015 Jun 17;2(2):69-88 [FREE Full text] [doi: 10.17083/ijsg.v2i2.66]
28. Saidin N, Abd Halim N, Yahaya N. A review of research on augmented reality in education: advantages and applications. Int Educ Stud 2015 Jun 25;8(13):1-8 [FREE Full text] [doi: 10.5539/ies.v8n13p1]
29. Ng K, Huang H, O'Malley C. Treasure codes: augmenting learning from physical museum exhibits through treasure hunting. Pers Ubiquit Comput 2018 Mar 13;22(4):739-750 [FREE Full text] [doi: 10.1007/s00779-018-1126-5]
30. Connolly T, Boyle E, MacArthur E, Hainey T, Boyle J. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012 Sep;59(2):661-686 [FREE Full text] [doi: 10.1016/j.compedu.2012.03.004]
31. Hainey T, Connolly T, Boyle E, Wilson A, Razak A. A systematic literature review of games-based learning empirical evidence in primary education. Comput Educ 2016 Nov;102:202-223 [FREE Full text] [doi: 10.1016/j.compedu.2016.09.001]
32. Wiemker M, Elumir E, Clare A. Escape room games: "Can you transform an unpleasant situation into a pleasant one?". 2015 Nov. URL: https://thecodex.ca/wp-content/uploads/2016/08/00511Wiemker-et-al-Paper-Escape-Room-Games.pdf [accessed 2022-09-20]
33. Nicholson S. Creating engaging escape rooms for the classroom. Child Educ 2018 Jan 11;94(1):44-49 [FREE Full text] [doi: 10.1080/00094056.2018.1420363]
34. Jiménez C, Arís N, Magreñán RA, Orcos L. Digital escape room, using Genially and a breakout to learn algebra at secondary education level in Spain. Educ Sci 2020;10:271 [FREE Full text] [doi: 10.3390/educsci10100271]
35. Williams P. Using escape room-like puzzles to teach undergraduate students effective and efficient group process skills. 2018 Presented at: ISEC 2018: IEEE Integrated STEM Education Conference; March 10, 2018; Princeton, NJ p. 254-257 URL: https://doi.org/10.1109/isecon.2018.8340495 [doi: 10.1109/isecon.2018.8340495]
36. Nicholson S, Cable L. Unlocking the Potential of Puzzle-Based Learning: Designing Escape Rooms and Games for the Classroom. Thousand Oaks, CA: SAGE Publications; 2021.
37. Borrego C, Fernández C, Blanes I, Robles S. Room escape at class: escape games activities to facilitate the motivation and learning in computer science. J Technol Sci Educ 2017 Jun 07;7(2):162 [FREE Full text] [doi: 10.3926/jotse.247]
38. Papadakis L, Stavrakis M. Experiential learning and STEM in modern education: incorporating educational escape rooms in parallel to classroom learning. 2020 Presented at: International Conference on Human-Computer Interaction; July 19-24, 2020; Copenhagen, Denmark p. 279-295 URL: https://doi.org/10.1007/978-3-030-50506-6_21 [doi: 10.1007/978-3-030-50506-6_21]
39. Sanchez E, Plumettaz-Sieber M. Teaching and learning with escape games from debriefing to institutionalization of knowledge. 2018 Presented at: GALA 2018: 7th International Conference on Games and Learning Alliance; December 5-7, 2018; Palermo, Italy p. 242-253 URL: https://doi.org/10.1007/978-3-030-11548-7_23 [doi: 10.1007/978-3-030-11548-7_23]
40. Adams V, Burger S, Crawford K, Setter R. Can you escape? Creating an escape room to facilitate active learning. J Nurses Prof Dev 2018;34(2):E1-E5 [FREE Full text] [doi: 10.1097/nnd.0000000000000433]
41. Vörös A, Sárközi Z. Physics escape room as an educational tool. 2017 Presented at: AIP Conference Proceedings; 2017; Bydgoszcz, Poland URL: https://doi.org/10.1063/1.5017455 [doi: 10.1063/1.5017455]
42. Dietrich N. Escape classroom: the Leblanc process—an educational "escape game". J Chem Educ 2018 Apr 20;95(6):996-999 [FREE Full text] [doi: 10.1021/acs.jchemed.7b00690]
43. Lopez-Pernas S, Gordillo A, Barra E, Quemada J. Comparing face-to-face and remote educational escape rooms for learning programming. IEEE Access 2021;9:59270-59285 [FREE Full text] [doi: 10.1109/access.2021.3073601]
44. Karageorgiou Z, Mavrommati E, Fotaris P. Escape room design as a game-based learning process for STEAM education. 2019 Presented at: ECGBL 2019: 12th European Conference on Game-Based Learning; 2019; Odense, Denmark p. 3-4 URL: https://doi.org/10.34190/gbl.19.190 [doi: 10.34190/gbl.19.190]
45. Gordillo A, Lopez-Fernandez D, Lopez-Pernas S, Quemada J. Evaluating an educational escape room conducted remotely for teaching software engineering. IEEE Access 2020;8:225032-225051 [FREE Full text] [doi: 10.1109/access.2020.3044380]
46. Vergne M, Smith J, Bowen R. Escape the (remote) classroom: an online escape room for remote learning. J Chem Educ 2020 Jul 24;97(9):2845-2848 [FREE Full text] [doi: 10.1021/acs.jchemed.0c00449]
47. Kroski E. What is a digital breakout game? Libr Technol Rep 2020;56(3):5-7 [FREE Full text]
48. Monnot M, Laborie S, Hébrard G, Dietrich N. New approaches to adapt escape game activities to large audience in chemical engineering: numeric supports and students' participation. Educ Chem Eng 2020 Jul;32:50-58 [FREE Full text] [doi: 10.1016/j.ece.2020.05.007]

XSL•FO
RenderX

49. Huang S, Kuo Y, Chen H. Applying digital escape rooms infused with science teaching in elementary school: learning performance, learning motivation, and problem-solving ability. Thinking Skills Creativity 2020 Sep;37:100681 [FREE Full text] [doi: 10.1016/j.tsc.2020.100681]
50. Al-Harthi A, Al-Asfour A. Meeting school leadership challenges of online education with design thinking during COVID-19. 2020 Presented at: econf 2020: Sixth International Conference on e-Learning; December 6-7, 2020; Sakheer, Bahrain p. 229-232 URL: https://doi.org/10.1109/econf51404.2020.9385423 [doi: 10.1109/econf51404.2020.9385423]
51. Melles G, de Vere I, Misic V. Socially responsible design: thinking beyond the triple bottom line to socially responsive and sustainable product design. CoDesign 2011 Sep;7(3-4):143-154 [FREE Full text] [doi: 10.1080/15710882.2011.630473]
52. Dam R, Siang T. What Is Design Thinking and Why Is It So Popular?. 2018. URL: https://www.interaction-design.org/literature/article/what-is-design-thinking-and-why-is-it-so-popular#:~:text=In%20essence%2C%20design%20thinking%3A,the%20assumptions%20and%20the%20implications [accessed 2022-09-20]
53. Plattner H. An Introduction to Design Thinking PROCESS GUIDE. URL: https://dschool-old.stanford.edu/sandbox/groups/designresources/wiki/36873/attachments/74b3d/ModeGuideBOOTCAMP2010L.pdf [accessed 2021-12-12]
54. Cross N. Design Thinking: Understanding How Designers Think and Work. Oxford; New York: Berg; 2011.
55. Taraldsen L, Haara F, Lysne M, Jensen P, Jenssen E. A review on use of escape rooms in education – touching the void. Educ Inq 2020 Dec 14;13(2):169-184 [FREE Full text] [doi: 10.1080/20004508.2020.1860284]
56. Korkmaz, Çakir R, Özden M. A validity and reliability study of the computational thinking scales (CTS). Comput Hum Behav 2017;72:558-569 [FREE Full text] [doi: 10.1016/j.chb.2017.01.005]
57. Videnovik M, Vlahu‐Gjorgievska E, Trajkovik V. To code or not to code: introducing coding in primary schools. Comput Appl Eng Educ 2020 Dec 03;29(5):1132-1145 [FREE Full text] [doi: 10.1002/cae.22369]
58. Genially. URL: https://genially.com/ [accessed 2021-01-06]
59. Ang J, Ng Y, Liew R. Physical and digital educational escape room for teaching chemical bonding. J Chem Educ 2020 Jul 29;97(9):2849-2856 [FREE Full text] [doi: 10.1021/acs.jchemed.0c00612]
60. Levy M, Hadar I, Aviv I. Agile-based education for teaching an Agile requirements engineering methodology for knowledge management. Sustainability 2021 Mar 06;13(5):2853 [FREE Full text] [doi: 10.3390/su13052853]
61. González-Zamar M, Abad-Segura E. Implications of virtual reality in arts education: research analysis in the context of higher education. Educ Sci 2020 Aug 29;10(9):225 [FREE Full text] [doi: 10.3390/educsci10090225]
62. Toda A, Valle P, Isotani S. The dark side of gamification: an overview of negative effects of gamification in education. In: Cristea A, Bittencourt I, Lima F, editors. Communications in Computer and Information Science, Vol 832. Cham: Springer; 2017:143-156.
63. Putz L, Hofbauer F, Treiblmaier H. Can gamification help to improve education? Findings from a longitudinal study. Comput Hum Behav 2020 Sep;110:106392 [FREE Full text] [doi: 10.1016/j.chb.2020.106392]

Edited by N Zary; submitted 16.07.21; peer-reviewed by Y Zidoun, S Downie, V Verma, M Eckert; comments to author 10.09.21; revised version received 30.10.21; accepted 12.08.22; published 26.09.22.

Please cite as:
Videnovik M, Vold T, Dimova G, Kiønig LV, Trajkovik V
Migration of an Escape Room–Style Educational Game to an Online Environment: Design Thinking Methodology
JMIR Serious Games 2022;10(3):e32095
URL: https://games.jmir.org/2022/3/e32095
doi:10.2196/32095
PMID:36155979

©Maja Videnovik, Tone Vold, Georgina Dimova, Linda Vibeke Kiønig, Vladimir Trajkovik. Originally published in JMIR Serious Games (https://games.jmir.org), 26.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Design of a Game-Based Training Environment to Enhance Mental Health Care Professionals' Skills in Using e–Mental Health: Multiple Methods User Requirements Analysis

Joyce Bierbooms[1], PhD; Milou A Feijt[2], MSc; Wijnand A IJsselsteijn[2], Prof Dr; Inge M B Bongers[1,3], Prof Dr

[1]Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[2]Human-Technology Interaction, Eindhoven University of Technology, Eindhoven, Netherlands

[3]GGz Eindhoven en de Kempen, Eindhoven, Netherlands

Corresponding Author:
Joyce Bierbooms, PhD
Tranzo
Tilburg School of Social and Behavioral Sciences
Tilburg University
P.O. box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 48354735
Email: J.J.P.A.Bierbooms@tilburguniversity.edu

Abstract

Background: A major factor hampering the adoption of technology in mental health care is a lack of knowledge and skills. Serious gaming offers a potentially effective strategy to enhance the skills needed through experiencing and learning-by-doing in a playful way. However, serious gaming solutions are not widely available for mental health care. Therefore, the development of a game-based training environment in mental health care was pursued in a design project. The first step in such a design project is to identify user requirements that should be met.

Objective: This study aims to deliver user requirements that inform the design of a game-based training environment for mental health care professionals. This environment aims to support mental health care professionals' knowledge and skill enhancement regarding the use of e–mental health (eMH); for example, video calling, mobile apps, web-based treatment modules, and techniques such as virtual or augmented reality.

Methods: We used an exploratory multiple methods design consisting of a web-based questionnaire, co-design sessions, and interviews. To ensure a good representation of the target user group, professionals from various disciplines within mental health care were included in the research. The multiple methods design facilitates a broad view of user needs and in-depth knowledge of specific design requirements. We describe the protocol for this research project in a protocol paper published in the *JMIR Research Protocols* in February 2021.

Results: The user requirements analysis revealed three types of users for the envisioned game-based training environment: mental health care professionals who want to learn about the basic possibilities of eMH, mental health care professionals who want to develop their eMH skills to the next level, and mental health care professionals who want to experiment with new technologies. This reflects the diversity of needs that were identified, as well as the need to develop a diversity of suitable scenarios in the environment. User requirements analysis shows that the focus of a training environment should be on increasing knowledge about the possibilities of eMH, focusing on experiencing the benefits in particular situations, and building confidence in using eMH in a therapeutic setting. This requires careful consideration of the suitable game characteristics.

Conclusions: Improvement of mental health care professionals' skills in eMH requires an environment that is user driven and flexible, and simultaneously incorporates contextual factors that are relevant for its implementation in practice. This user requirements analysis contributes to the understanding of the issues that should be considered in the development of a game-based training environment. This shows that there are multiple and diverse learning needs among mental health care professionals. Various client populations, services, and situations demand various options for training.

International Registered Report Identifier (IRRID): RR2-10.2196/18815

XSL•FO
RenderX

(JMIR Serious Games 2022;10(3):e34700) doi:10.2196/34700

KEYWORDS

serious gaming; e–mental health; mental health care; skill development; game design; user requirements; mobile phone

Introduction

Background

Although the rise of new technologies has offered numerous possibilities for new methods and means of treatment in mental health care, the uptake of such technologies by mental health care professionals is rather slow [1-3]. With such technologies, we refer to mental health services that are provided or supported by digital tools; for example, video calling, mobile apps for self-tracking or self-management, web-based treatment modules, and techniques such as virtual or augmented reality [4-6], hereafter referred to as e–mental health (eMH). One important factor hampering the adoption of eMH by mental health care professionals is a lack of knowledge and skills of mental health care professionals to effectively find and use web-based technologies and their corresponding need to gain more experience in this [7-10]. This knowledge and skill gap concerns different types of knowledge and skills regarding the use of eMH [8,11]. For example, a lack of skills may exist regarding the ability to properly use technological devices or regarding knowledge about the availability of eMH tools that can be applied for different therapies. Another example is the perceived difficulties in establishing rapport in remote forms of communication [8,11].

Skill Enhancement Through Game-Based Training

Multiple training concepts have been developed in mental health care to enhance professionals' skills in using eMH, such as classical instructions, individual e-learning, and demonstrations. However, they have not reached the point at which this has led to a substantial expansion of knowledge and skills regarding eMH [12]. The reasons for this are the lack of interactivity and experiential learning in these concepts [13]. Therefore, a new potential strategy to address this issue may be to offer mental health care professionals training possibilities in a game-based environment, where skill enhancement is achieved through experiencing and learning-by-doing in a playful way [13]. Multiple positive outcomes of game-based learning have already been demonstrated in other areas of business [14], and positive effects have been reported in several studies on serious gaming to train health care professionals [13]. A major advantage of serious gaming lies in the fact that it offers a hands-on learning experience rather than just reading or hearing instructions [14,15]. It also offers a unique combination of education and fun [16]. Furthermore, serious games have demonstrated various learning outcomes (eg, cognitive skills, motor skills, affective learning outcomes, and communicative learning outcomes) that are more difficult to achieve through traditional learning methods [17]. These are strong arguments to believe that serious gaming is a potentially effective way to develop mental health care professionals' skills in using eMH in an experience-based way. Moreover, it would be difficult, expensive, and unethical to experiment with real clients in this way [18,19]. Looking at the current literature on serious gaming, it can be noticed that such training possibilities have not yet been specifically designed for mental health care. Therefore, we intend to design a game-based training environment specifically for mental health care professionals in which they can explore eMH, practice with different technological tools, and enhance their knowledge and skills in eMH. Finally, this could stimulate the increasing use of eMH in daily practice.

To establish a game-based training environment that meets the specific needs of mental health care professionals, they should be involved from the start to inform designers of their needs and preferences. In this way, the design will reflect their actual work situations, the problems they encounter in these situations, and the needs and preferences regarding ways to tackle these problems [20-23]. The first step in such a design project is to identify the user requirements that the training environment should satisfy [24-27]. This means identifying and describing end users' needs in terms of their characteristics, learning needs, and learning preferences. This information is used to develop scenarios that reflect the real-life situations of these end users and draw up a list of technical and organizational requirements that this game-based training environment should meet [12]. In this paper, we describe the results of the user requirements analysis that was conducted in light of the design of a game-based training environment for mental health care professionals in the Netherlands. This analysis aimed to gain a better understanding of mental health care professionals' attitudes, skills, and ambitions regarding the use of eMH.

Objectives

This study aimed to deliver user requirements that inform the design of a game-based training environment for mental health care professionals to enhance their knowledge and skills regarding the use of eMH [12]. The user requirements consist of (1) an elaborate analysis of end users' (ie, mental health care professionals') needs, (2) a description of possible scenarios regarding the content of a game-based training environment, and (3) the technical and organizational requirements that are expected to be critical for the successful implementation of a game-based training environment.

Methods

Overview

In this section, we provide a brief summary of the methods used, which are described in detail in a protocol study [12]. To extract user requirements, we used an exploratory multiple methods design consisting of a questionnaire, co-design sessions, and interviews. We used a web-based questionnaire (N=432) administered to various mental health care organizations to gather data on end users' needs in terms of mental health care professionals' characteristics, attitudes, and skill levels regarding eMH. A co-design session was organized with 9 mental health care professionals to elaborate on these characteristics, attitudes, and skill levels and to gain the first ideas about mental health

XSL•FO
RenderX

care professionals' learning preferences in a game-based training environment. For the data collection that informed the scenarios, another 2 co-design sessions were held, each with 10 participants, including mental health care professionals, innovations experts, developers, and researchers. Finally, we conducted 17 interviews with mental health care professionals. These interviews were aimed at learning about the preferences of mental health care professionals regarding training in eMH and the technical and organizational prerequisites for the envisioned game-based training environment. Textbox 1 presents an overview of the methods, describing the aim of each method and the participants involved.

XSL•FO
RenderX

Textbox 1. Overview of the aim and participants of each study method.

Web-based questionnaire

- Aim
  - Description of the characteristics of the user group
  - Perceived problems related to e–mental health skills
- Participants
  - N=432
  - Male: 144 (33%)
  - Female: 288 (67%)
  - Average age: 41
  - Professions
    - Nursing professionals (32.4%, including specialized nursing)
    - Psychologists (27.5%, also including specializations)
    - Social workers (17.1%)
    - Other (23%; for example, medical specialists, case managers, activity workers, and supporting staff)

Co-design session 1

- Aim
  - Work context
  - Perceived problems related to e–mental health skills
  - Users' preferences regarding a game-based training environment
- Participants
  - N=9
  - Male: 2 (22%)
  - Female: 7 (78%)
  - Professions
    - Medical specialist (n=1, 11%)
    - Psychologists (n=2, 22%)
    - Nursing professionals or specialized nurses (n=4, 44%)
    - Social workers (n=2, 22%)

Co-design session 2

- Aim
  - Scenario development
- Participants
  - N=10
  - Male: 5 (50%)
  - Female: 5 (50%)
  - Professions
    - Mental health care professionals (n=3, 30%)
    - Game developers (n=3, 30%)
    - Innovation experts (n=2, 20%)

XSL•FO
RenderX

- Researchers (n=2, 20%)

Co-design session 3

- Aim
  - Scenario development
- Participants
  - N=10
  - Male: 5 (50%)
  - Female: 5 (50%)
  - Professions
    - Mental health care professionals (n=3, 30%)
    - Game developers (n=3, 30%)
    - Innovation experts (n=2, 20%)
    - Researchers (n=2, 20%)

Interviews

- Aim
  - Users' preferences regarding a game-based training environment
  - Technical and organizational prerequisites
- Participants
  - N=17
  - Male: 4 (24%)
  - Female: 13 (77%)
  - Professions
    - Medical specialist (n=1, 6%)
    - Psychologists (n=3, 18%)
    - (Specialized) nursing professionals (n=4, 24%)
    - Social workers (n=3, 18%)
    - Innovation staff (n=4, 24%)
    - Management (n=2, 12%)

To ensure a good representation of the target user group for each data collection method, professionals from various disciplines involved in the direct care delivery process within mental health care were included in the research. Professionals of 5 large integrated mental health care organizations providing mental health care to clients with complex mental health problems participated in the questionnaire. Co-design sessions and interviews were conducted at one of these organizations (Stichting Geestelijke Gezondheidszorg Eindhoven en de Kempen [GGzE]) in the southern part of the Netherlands. The multiple methods design combines a broad view of user needs (through the web-based questionnaire) and in-depth knowledge of specific design requirements (through interviews and co-design). The questionnaire data were used to analyze, using descriptive statistics, which subgroups could be defined among mental health care professionals as potential users of eMH. In addition, the data were used to define the specific needs for each subgroup. The co-design sessions and interviews were analyzed using thematic coding and provided in-depth information on the requirements for a game-based training environment based on specific end user's needs. The data were triangulated as described in the protocol paper [12].

Ethics Approval

This study was approved by the ethical review board of Tilburg University (reference number: EC-2018.15) and the internal scientific review board of GGzE (reference number: MM/2019004).

XSL•FO
RenderX

Results

Overview

Following the approach described in the Methods section, this section presents the results of the user requirements analysis in the design process of a game-based training environment. First, we clarify the end users' needs regarding a game-based training environment by providing: (1) a description of the characteristics of the user group (ie, mental health care professionals), (2) their work context and practices, (3) their perceived problems regarding the use of eMH, and (4) their preferences related to a game-based training environment. Subsequently, we describe the scenarios by describing how the added value of eMH tools can be effectuated in specific situations, combined with relevant workflows and processes, the use of specific eMH tools in different situations, and decisive moments regarding the use of eMH. Finally, we present the results of the study aimed at identifying the technical and organizational conditions that are needed to enable the use of the anticipated game-based training environment. This means that design- and content-related conditions as well as organizational stimuli that are required to launch a game-based training environment in practice were explored.

End Users' Needs

Description of the Characteristics of the User Group

On the basis of the survey data from mental health care professionals in the Netherlands, a number of general characteristics of the sample can be described. Most mental health care professionals were female (288/432, 66.8%), with an average age of 41.3 (SD 12.1) years, and many of them had been working as health care professionals for a relatively long period (average of 16.3 years, SD 11.4). Mental health care professionals were highly educated: 49.3% (213/432) of them have received higher vocational education and 38.7% (167/432) have an academic degree. The main professions that were found among the respondents of the survey were nursing professionals (140/432, 32.4%, including specialized nursing), psychologists (119/432, 27.5%, also including specializations), and social workers (74/432, 17.1%). Most mental health care professionals (278/432, 64.4%) indicated that they had not received specific eMH training.

To identify the extent to which mental health care professionals in the Netherlands have adopted eMH in their daily practice, the survey included a number of questions related to the adoption of eMH (Table 1). Only 26.3% (114/432) of mental health care professionals were active or innovative users of eMH.

Table 1. Levels of adoption of e–mental health (eMH) by mental health care professionals in the Netherlands (N=432).

Level of adoption of eMH	Values, n (%)
No use ("I do not really want to start using it")	11 (25)
Minimal use ("It is not part of my daily routine")	112 (26)
Passive use ("I use what is readily available")	195 (45.2)
Active use ("I am exploring more possibilities")	86 (19.8)
Innovative use ("I am going to build upon my new idea")	28 (6.5)

Work Context

The first co-design session aimed to gather data to identify the work context of mental health care professionals. The data showed that mental health care professionals in the co-design session work with a client population with multiple complex mental health problems, including clients in a crisis situation, which means they work in a specialized mental health care setting. They worked in a clinical or ambulatory setting, the latter being the setting in which most of the respondents worked. Most ambulant treatment trajectories aimed to support clients during their recovery process. This means working toward a situation in which people are able to manage their own mental well-being and know how to cope with difficult situations to improve their quality of life. In many cases, mental health care professionals are also in contact with the family and friends of their clients, as well as organizations such as schools and financing institutions that are often involved in the treatment trajectories. Mental health care professionals state that it may be opportune to offer eMH options at some point during a treatment trajectory, for example, to keep contact with a client's network or monitor a client's functioning between treatment meetings.

Perceived Problems Related to eMH Skills

As mentioned in the Introduction section, one of the key problems for the large-scale adoption of eMH among mental health care professionals is the perceived lack of skills [7-10]. Insights from the survey provided an indication of how mental health care professionals perceive their skill levels regarding the use of eMH. Table 2 shows that a large percentage of mental health care professionals reported insufficient knowledge about the availability of eMH tools or lack knowledge on how to apply these tools. Of the people who are knowledgeable about how to use eMH to some extent, a large proportion only know the basics of a small number of tools.

XSL•FO
RenderX

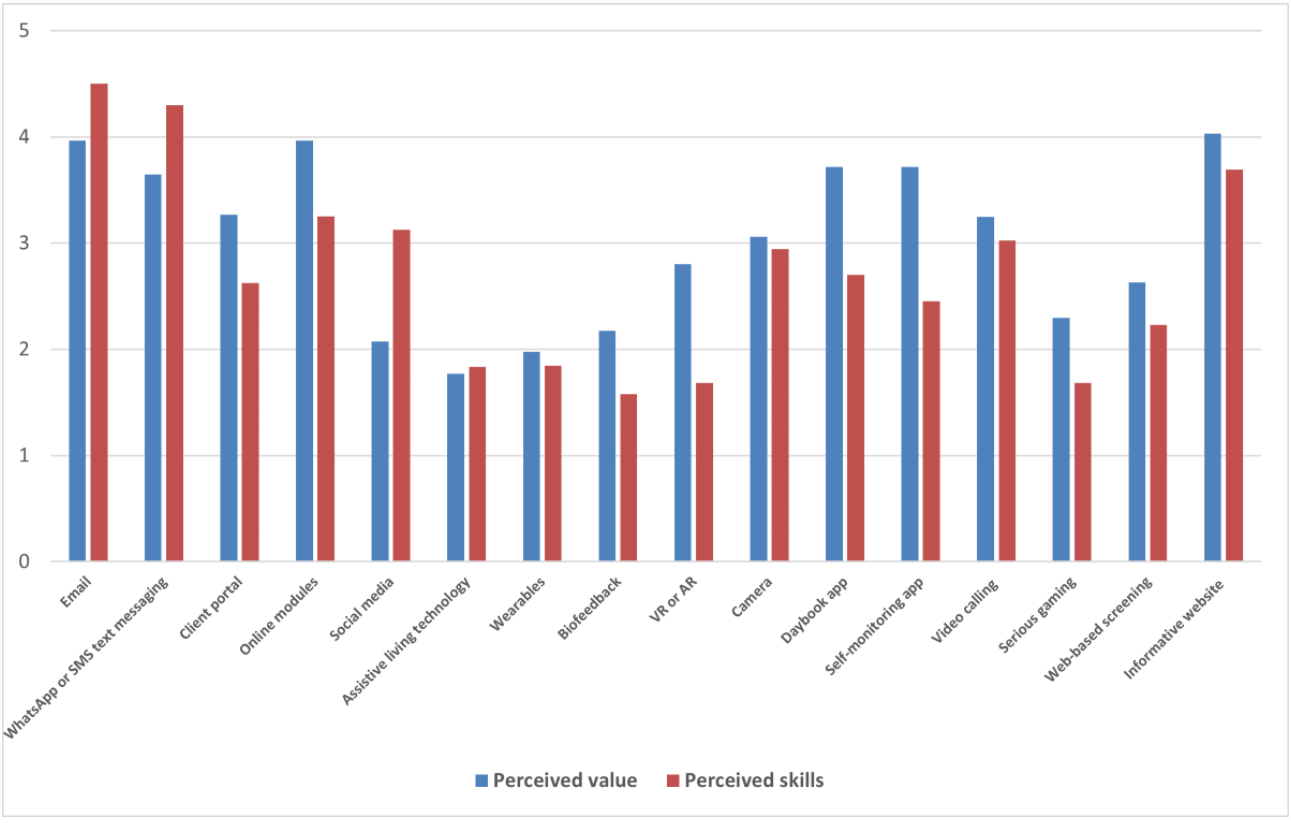
Table 2. Self-perceived skill levels regarding the use of e–mental health (eMH) by mental health care professionals (N=432).

Self-perceived skill level	Values, n (%)
I have insufficient knowledge of what is available when it comes to eMH	104 (24.1)
I have an image of what eMH tools are available, but I do not know how to use these	74 (17.1)
I know how to use the basic functionalities of eMH	133 (30.8)
I am capable to explore the possibilities of eMH and how to use different tools	97 (22.5)
I feel very competent regarding many different eMH tools and I know how to share this expertise with others	24 (5.6)

When looking further into these skills, both the survey and the first co-design session revealed that different skill types play a role in using eMH. Most respondents indicated that they felt they had sufficient digital skills but lacked knowledge on how to communicate in a web-based contact with their clients. Moreover, the feeling of being unable to establish sufficient empathic understanding in a web-based conversation is an important issue for mental health care professionals. The types of skills that mental health care professionals feel they need to practice most, differ depending on their level of adoption of eMH. For example, the co-design session revealed that people who qualify themselves as minimal to passive users often want to find out more about what is actually available regarding eMH. People who are already active users want to take it a step further and want to discover new possibilities on how to create an empathic interaction.

The perceived skill levels of mental health care professionals also differ among the types of eMH tools that are available (Figure 1). For relatively new and innovative tools, such as virtual reality and biofeedback, there is a pronounced discrepancy between the perceived value and the perceived skill level with regard to using such a tool is rather high. That is, the perceived value of these tools is relatively high compared with the perceived skill level. This is also the case for tools that involve self-monitoring or a web-based diary. Knowing these discrepancies provides input for designers regarding which eMH tools would fit in a game-based training environment for skill enhancement in eMH. That is, these results help to identify eMH tools that are perceived to be valuable but where lack of skills might hamper their uptake and use in clinical practice. Therefore, these tools are particularly suitable for addressing training environments. In the co-design session, mental health care professionals were provided with more detailed information about the possibilities offered by various eMH tools. After these explanations and examples, the participants in the co-design session perceived the tool to have more value and showed an increased willingness to learn more about these tools compared with their opinions before the session.

Figure 1. Perceived value and perceived skills of e–mental health tools. AR: augmented reality; VR: virtual reality.



XSL•FO
RenderX

Users' Preferences Regarding a Game-Based Training Environment

User preferences for the design of a game-based training environment cover both the general features of the environment and the ways in which the content is being delivered. These preferences were gathered during the interviews and the first co-design session. First, mental health professionals indicated that there should be a clear-cut purpose and that tangible advantages for the user should become instantly clear when engaging in the game environment, as illustrated by the following quote:

> *In terms of what is in it for me? It may sound stupid, but work pressure is very high, so I don't have a lot of time for these kinds of things. ... So when I can see that it has a clear purpose and that I will benefit from it in the future; then I would participate.* [Respondent 1, psychologist]

The game environment should be easily accessible and contain customizable game elements for users with different backgrounds (eg, different levels of adoption, skills, and knowledge) and subsequent training needs. Furthermore, participants indicated that time investment should be done as efficiently as possible, including a smooth transfer of acquired knowledge and skills to their daily practice. This means that it is important to know the actual working situations of mental health care professionals to simulate this as close to reality as possible. Related to this, mental health care professionals stress the need for a game-based training environment in which they can experience the benefits of eMH and practice by building confidence regarding its use in therapeutic practice. A participant stated the following:

> *If it is offered and I am convinced of its benefits, then I will use it. It should eventually benefit the treatment of my clients, if I can see that it adds to that, then I will more easily use it.* [Respondent 5, social worker]

Finally, there should be a certain amount of learner control embedded in the game solution, implying that the system should be sufficiently flexible to address different types of needs. This means training for different skill levels, types of skills, flows, and pathways that can be constructed by the users themselves. To satisfy these requirements, it is important to involve end users in the content and design of the environment, as stated by mental health care professionals in this study.

In addition to these preferences regarding the design of the environment, there are also several considerations that mental health care professionals have expressed as to how the content is delivered. The respondents indicated that they were looking for an engaging and playful way of learning with well-defined learning goals. This can be realized by creating a system in which students receive constant and quick feedback and are challenged to take the next step in their learning process. Moreover, the game should be a complete learning experience, with a clear learning goal that is evaluated after playing the game. This is essential for mental health care professionals to transfer from the learning experience to their therapeutic practice. This should also provoke an inductive way of learning. That is, it should support a process of discovering more general rules from specific examples, which enhances the problem-solving capabilities of users. A final need that was mentioned multiple times was the possibility of learning together with colleagues or peers. Making eMH training a team activity would, according to mental health care professionals, be much more effective and engaging than individual training in specific skills. A participant stated the following:

> *So I think you should make it a group activity and create a good context for using it. … With my team I would take the challenge, but not on my own in my room playing a single player game.* [Respondent 13, psychologist]

User Groups for the Design of a Game-Based Training Environment

Assembling the results on the end users' needs, 3 main user groups of a game-based training environment can be described. First, a group of mental health care professionals, mainly highly educated nurses, who wanted to learn about the basic possibilities of eMH without putting too much effort and time into it. Training in eMH needs to bring forward positive experiences which in turn could lead to the incorporation of eMH in daily routines. Second, a group of mental health care professionals, nurses, and psychologists, who aimed to bring their eMH skills a step further. These mental health care professionals had already incorporated the basics of eMH in their daily routines but wished to know more about different eMH tools. This group wanted to learn about, for example, the possibilities of dealing with a lack of nonverbal communication in eMH or handling boundaries toward clients. Third, a group of mental health care professionals from various professions, who sought to learn something different. This group of mental health care professionals was eager to try new tools and did not want to be confronted with basic exercises in a game-based training environment. They sought support from team members and wanted to know how to transfer positive experiences when using eMH to colleagues. These 3 main user types reflect the diversity of needs we identified and need to be incorporated into our design through a diversity of suitable scenarios or inherent layering of complexity within a training environment.

Scenario Development

At the beginning of the Results section, we explained scenario development as an identification of the goals, tasks, actions, and decisions mental health care professionals encounter regarding the use of eMH in daily practice. In this section, we describe the results for these scenarios found in two co-design sessions (sessions 2 and 3; Textbox 1) that were specifically aimed at scenario development.

Goals: Purpose of eMH

In the second co-design session, mental health care practitioners were asked to reflect on the purpose of eMH and the steps they deemed important to fulfill this purpose. This led to the description of a catchphrase with regard to the use of eMH in the future: mental health care professionals are familiar and skilled in using innovative treatment solutions, establishing an even better match between mental health services and clients' needs. The added value of eMH should become clear by

XSL•FO
RenderX

demonstrating that web-based treatment can also provide a rich context; for example, it provides mental health care professionals with observable information about someone's living environment and circumstances. The co-design session also clarified that skill enhancement should enable mental health care professionals to make adequate estimations of the use of different eMH tools in specific situations. Participants stated that knowledge and skills to use eMH should be part of every mental health care professional's skill set and should be a mandatory competence in every mental health care organization. According to mental health care professionals in the co-design session, eMH has the potential to contribute to much better access to mental health care for many people. For example, a general practitioner can refer clients directly to web-based mental health care without waiting lists.

Tasks: Clients' Characteristics and Therapeutic Situations

According to the survey results and mental health care professionals in the third co-design session, interactions with clients can be aimed at clarifying the problem, finding solutions together, and helping the client handle problematic situations on their own, in other words, knowing how to recognize and cope with different signals with the goal of increasing the client's symptomatic, personal, and social recovery. In addition, psychoeducation, relapse prevention, and rehabilitation are mentioned as possible therapeutic situations. Whether clients value face-to-face contact with therapists differs among clients. A number of mental health care professionals have reported that some clients would prefer eMH over face-to-face contact, which would even stimulate their engagement and adherence to treatment. This is in line with how the participants in the first co-design session described their work context.

Actions: Use of Specific eMH Tools

The range of currently available eMH tools in the practice of mental health care professionals comprises a rather diverse set of tools in terms of maturity and even more so for the level of advancement in training for these tools. The eMH tools that were pointed out in the second and third co-design sessions as worthy of further exploration in a potential game-based training environment are a web-based platform for psychological treatment (consisting of tools such as video calling, web-based modules, and a message function), WhatsApp, self-help programs, web-based group courses, a client portal, web-based screening, virtual reality, and biofeedback (including wearables). This finding is in line with the results of a survey that also elucidated these types of eMH tools. It was emphasized that for mental health care professionals to experience the benefits of using eMH when working with a client, it is important that the type of eMH fits the particular characteristics and needs of the client. In addition, mental health care professionals have stressed the importance of offering eMH tools that are ready to use in practice. Regarding eMH tools that are still in their infancy or experimental stage, a specific group of pioneers can be approached. However, these tools will be less appealing to most mental health care professionals who are still in search of gaining experience with the basic principles of using eMH in general.

Decisions: Choice to Use eMH Based on Anticipated Consequences

The nature and complexity of psychological disorders, the client's age, level of digital literacy, intelligence, and the devices available to the client are factors that are often mentioned by mental health care professionals as having an important influence on the use of eMH in practice. The specific desires and needs of clients regarding the use of eMH tools are something a professional and client need to explore and discover together. In the co-design sessions on scenario development, an important aspect mentioned for the continued use of eMH is that professionals experience is a clear consequence or effect related to the use of eMH in their work. Although there are various treatment trajectories one can think of and various client groups that mental health care professionals see, it is important to create multiple scenarios with different outcomes (ie, different decisive moments to use eMH and perceive the benefits). Moreover, participants in the co-design sessions indicated that it would be valuable if there were several points in a scenario in which participants had multiple options to choose, leading to different experiences of eMH. There needs to be a clear conclusion to the story; for example, that the client is finding a solution to their problem. This should be related to the choices made in the game, and in particular, the choices with regard to using eMH. Providing this conclusion from the perspective of the client within a scenario would, according to the participants of the co-design sessions, significantly contribute to the effectiveness of a game-based training environment.

Technical and Organizational Prerequisites

In the development of a game-based training environment, it is important to consider the technical and organizational conditions (ie, the prerequisites that need to be met when delivering the final solution). These are crucial for the successful implementation of an envisioned environment within a mental health care organization. Several criteria were identified based on these interviews. A first key aspect elicited in the interviews was the technical support available from their organization when using the environment. This is especially true when offering a technical solution to people who still need to acquire the skills necessary to make more use of technological tools in mental health care in general. Technical issues must be resolved immediately to prevent the risk of negative experiences. A respondent stated this in a straightforward manner:

> *If it doesn't work, I am out.* [Respondent 3, psychologist]

The second aspect is the perceived added value of the training. Mental health care professionals indicated that they had to be convinced of the benefits of game-based training to feel motivated to engage in it. It should therefore be stressed beforehand that using the game-based training environment will eventually lead to better job performance and better treatment for mental health clients. This added value should be communicated, for example, by peer users or educational experts within the organization. Another important factor is management's push to engage in training. A significant number of mental health care professionals feel that they need some

XSL•FO
RenderX

form of obligation to take that step to engage in a game-based training environment. A participant formulated the following:

> *What should be facilitated by management is a clear assignment. ... That is something that I really miss now that we are a self-managing team. I want to know what space I have and what expectations management has. ... Beginning with defining clear goals for the next year, that is the job of a manager.* [Respondent 4, specialized nursing professional]

Experiencing a certain lack of skills, and the consequences of that, will trigger a sense of urgency by mental health care professionals. However, the way in which the environment is introduced is crucial. As soon as professionals feel that the management push is solely driven by the goal of saving money and not because professionals or their clients benefit from it, this pressure will elicit a strong feeling of resistance and hamper the adoption of the environment. Therefore, mental health care professionals feel that it is necessary for management to emphasize the added value it can provide. Management support in practical terms is equally important: to enable the effective use of the game-based training environment, time and resources should be made available for mental health care professionals to engage in it. Finally, several professionals mentioned low visibility and awareness of eMH during daily practice as an important influence on the adoption of eMH. Visibility and awareness are also important to consider when developing a game-based training environment. According to the interviewees, external triggers, standard procedures, and good embeddedness in organizational processes are crucial for the adoption of a game-based training environment. A respondent described this as follows:

> *I think it is an important aspect of the human resource strategy. As an organization you should focus on keeping your employees engaged and loyal. But also embedding this in standard procedures, as in: this is how we work and what we expect of our employees.* [Respondent 12, manager]

Discussion

Principal Findings

In this study, we aimed to identify the user requirements that inform the design of a game-based training environment for mental health care professionals to enhance their knowledge and skills regarding the use of eMH. This has led to specific guidelines for the design of a game-based training environment and has generated multiple ideas regarding its potential content and shape. In addition, the results of this study are useful for gaining a better understanding of mental health care professionals' attitudes, skills, and ambitions regarding the use of eMH. This can also be used to develop policy interventions to enhance the adoption and implementation of eMH among mental health care professionals and innovation in mental health care.

We identified that the adoption of eMH by mental health care professionals is low and that only 26.2% (113/432) of professionals claim to be active or innovative users of eMH. It is important to note that the survey that generated these results was conducted before the COVID-19 pandemic that started in March 2020, which led to an immediate large-scale (forced) use of eMH, particularly video calling [28]. However, this is not per definition equal to the sustainable adoption of eMH by professionals in the so-called *new normal* in mental health care [29]. In a recent study conducted during the first wave of the COVID-19 crisis, mental health care professionals indicated that the continued use of eMH largely depends on the experienced possibilities of eMH during the COVID-19 period and the context (ie, type of treatment, type of clients and their preferences, and clients' home situations) in which eMH is applied [29]. At this point in time, sustainable uptake and integration of eMH in the daily practice of mental health care professionals seems to have not yet taken place, despite the fact that many mental health care professionals have gained experience in using eMH under circumstances that require remote treatment [30]. This may be caused by a persistent lack of knowledge and skills regarding the possibilities of eMH or a general conception among mental health care professionals that eMH equals video calling and using web-based modules [31]. There is still relatively little knowledge about different types of eMH [31], and there is still a world to be discovered regarding the possibilities of, for example, virtual reality, biofeedback, and smartphone apps. Furthermore, from the existing literature, we have learned that knowledge and skills consist of different elements; for example, skills to gather information through remote therapy or to create an emphatic interaction [8,13]. This is consistent with the results of our user requirements analysis, which shows that the focus of a training environment for skill enhancement should be aimed at increasing knowledge about the possibilities of eMH in a broader sense, experiencing the benefits in particular situations, and building confidence in using eMH in a therapeutic setting. This requires careful consideration of the game characteristics (eg, player interaction, feedback, and competition) that are suitable for this aim [32,33]. In addition, we determined that there are differences in the characteristics, environment, perceived problems, and preferences of mental health care professionals that should be considered. This means that it is important to address the diversity that exists within the user group and the different requirements that this brings forward regarding a game-based training environment and the way in which it is implemented. This is also mentioned in the literature, where it is discussed that to empathize with users, designers have to know the different user groups [34].

Research into scenarios was aimed at identifying different situations that represent the daily practice of mental health care professionals in which eMH could play a role. For developers of a game-based training environment, it is important that a varied user group can identify the scenarios that are incorporated in the environment [35,36]. For example, mental health care professionals consider various eMH tools valuable. These tools differ in terms of maturity and skills required to being able to use them. Therefore, there is a need to match users' preferences to the maturity of the tools but also to align the content of a game-based training environment with the required skills and knowledge that users (feel they) are lacking. The results also showed that an important consideration for mental health care

XSL•FO
RenderX

professionals to continue using eMH is that a clear and positive outcome is experienced. This means that participants should be allowed to experience the consequence of a certain choice within the game [37], which in this case is the result of using eMH and possible added value. The various treatment trajectories with various client groups indicated that adaptive multilayered scenarios are required to address the practice of the varied group of mental health care professionals.

When developing a training environment, it is crucial to consider how this system will be implemented, that is, the context in which it will be offered and the technical and organizational conditions required [38]. For example, it could be offered by a mental health care organization as an option to support their employees in applying eMH, which could be part of separate vocational training (either optional or mandatory) or incorporated in the curriculum for future mental health care practitioners. Currently, to the best of our knowledge, there are no specific guidelines for training on how to integrate eMH into daily practice; in general, it is not an integral part of the education of mental health care professionals, although several organizations and institutions have started to offer courses on eMH. It is also important to have a clear vision of the demands a mental health care organization sets regarding the intention and skills to use of eMH by its employees and the capacity the organization has to support its employees in meeting these demands. There are several clear expectations that mental health care organizations can have regarding the use of eMH by mental health care professionals (eg, knowing a broad range of possible eMH tools and being able to apply them, a positive attitude toward eMH, and conforming to a new normality). This requires a reconsideration of the ambitions, a smart formulation of the goals the organization has regarding the use of eMH (eg, in the form of formal job requirements) and a plan to support these professionals in acquiring and retaining the active use of the digital possibilities of this era, as well as in mental health care.

Limitations

The data collection of this user requirements analysis took place in the period before the global pandemic, which may have influenced mental health care professionals' knowledge and skills regarding the availability of eMH tools in comparison with the results that were collected in this study. However, the use of eMH during the COVID-19 pandemic largely came down to video calling, which means there are many other eMH tools still to be discovered for mental health care professionals [28]. In addition, as becomes clear from current research, there remains a pressing need to invest in training and education of mental health care professionals to effectively use eMH in client-therapist interactions [39].

Another aspect that calls for further exploration is the development of scenarios that more prominently include the client perspective. It should be considered in future research that clients are involved in such design and development processes.

Conclusions

In this study, we aimed to deliver user requirements to inform the design of a game-based training environment for eMH skill enhancement. To enable a significant improvement in mental health care professionals' eMH skills, it is important that such an environment is user driven and flexible and simultaneously incorporates the contextual factors that are relevant for its implementation in practice. This user requirements analysis contributes to the understanding of the issues that should be considered in the development of a game-based training environment by showing that "the" mental health care professional does not exist and that a variety of client populations, services, and situations demand a variety of options for training (eg, different difficulty levels and multiple story lines). We used this knowledge for the iterative design process of the serious gaming concept in this project. We designed an eMH escape room with 2 different storylines and explored different eMH tools in different ways. Mental health care professionals were also involved in the design process. These findings can also be of value for other eMH development or implementation projects or for other projects aimed at designing novel learning tools for mental health care professionals in general. Finally, and perhaps even more importantly, this study has provided a very broad and in-depth understanding of what facilitates and hampers the large-scale adoption of eMH in mental health care in general. Therefore, we feel that the results of this study could not only be used as requirements for the design of (technical) solutions for skill enhancement but also as input for policy interventions to stimulate the use of eMH. This can enhance sustainable change regarding the adoption of eMH in mental health care.

Acknowledgments

The authors would like to thank all the mental health care professionals and experts from other disciplines in the field that participated in this research. The authors thank Samira Veninga, MSc, for conducting the interviews as part of her Master's thesis and Dr Karin van den Berkmortel for facilitating the first co-design session.

This work was supported by Dutch Research Council grant 055.16.141; project title Psychosocial Games Supporting Mental Health Professionals in Stress Reduction and Empathic Interactions in Remote/Online Psychotherapy, under the Serious Games for Professional Skills Program.

Authors' Contributions

JB and MAF administered the questionnaire and analyzed the results. JB coded and analyzed the interview data and data from the co-design sessions. JB composed the initial draft of the manuscript, after which all authors contributed to multiple revisions. All authors have agreed to the final submitted version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Kilbourne AM. E-health and the transformation of mental health care. Psychiatr Serv 2012 Nov;63(11):1059. [doi: 10.1176/appi.ps.631103] [Medline: 23117506]
2. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark S, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
3. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
4. Lal S, Adair C. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
5. McGinty K, Saeed S, Simmons S, Yildirim Y. Telepsychiatry and e-mental health services: potential for improving access to mental health care. Psychiatr Q 2006;77(4):335-342. [doi: 10.1007/s11126-006-9019-6] [Medline: 16927161]
6. van der Krieke L, Wunderink L, Emerencia AC, de Jonge P, Sytema S. E-mental health self-management for psychotic disorders: state of the art and future perspectives. Psychiatr Serv 2014 Jan 01;65(1):33-49. [doi: 10.1176/appi.ps.201300050] [Medline: 24129842]
7. Carper MM, McHugh RK, Barlow DH. The dissemination of computer-based psychological treatment: a preliminary analysis of patient and clinician perceptions. Adm Policy Ment Health 2013 Mar;40(2):87-95. [doi: 10.1007/s10488-011-0377-5] [Medline: 22001968]
8. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
9. Donovan C, Poole C, Boyes N, Redgate J, March S. Australian mental health worker attitudes towards cCBT: what is the role of knowledge? Are there differences? Can we change them? Internet Interv 2015 Nov;2(4):372-381. [doi: 10.1016/j.invent.2015.09.001]
10. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
11. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
12. Bierbooms JJ, Sluis-Thiescheffer WR, Feijt MA, IJsselsteijn WA, Bongers IM. Design of a game-based training environment to enhance health care professionals' e-mental health skills: protocol for a user requirements analysis. JMIR Res Protoc 2021 Mar 17;10(2):e18815 [FREE Full text] [doi: 10.2196/18815] [Medline: 33595453]
13. Wang R, DeMaria S, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Mar;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
14. Squire K, Jenkins H. Harnessing the power of games in education. Insight 2003;3(1):5-33.
15. Kolb D, Boyatzis R, Mainemelis C. Experiential learning theory: previous research and new directions. In: Perspectives on Thinking, Learning, and Cognitive Styles. New York: Routledge; 2001.
16. Prensky M. Fun, play and games: what makes games engaging. In: Digital Game-Based Learning. New York: McGraw-Hill; 2001.
17. Connolly T, Boyle E, MacArthur E, Hainey T, Boyle J. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012 Sep;59(2):661-686 [FREE Full text] [doi: 10.1016/j.compedu.2012.03.004]
18. Reeves B, Read L. Total Engagement Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete. Boston, Massachusetts, United States: Harvard Business Press; 2009.
19. Serious Games Mechanisms and Effects. New York: Routledge; 2009.
20. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction. Thousand Oaks, California: SAGE Publications; 2004.
21. Nielsen J. The usability engineering life cycle. Computer 1992 Mar;25(3):12-22. [doi: 10.1109/2.121503]
22. Norman D. The Design Of Everyday Things: Revised And Expanded Edition. New York, United States: Basic Books; 2013.
23. User-centered design in games. In: The Human–Computer Interaction Handbook. Boca Raton, Florida, United States: CRC Press; 2012.
24. Bate P, Robert G. Bringing User Experience to Healthcare Improvement The Concepts, Methods and Practices of Experience-based Design. London: Radcliffe Publishing; 2007.
25. Maguire M, Bevan N. User requirements analysis. In: Usability. Boston, MA: Springer; 2002.
26. Wiegers K, Beatty J. Software Requirements, 3rd Edition. London, United Kingdom: Pearson Education; 2014.

XSL•FO
RenderX

27. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI07: CHI Conference on Human Factors in Computing Systems; Apr 28 -May 3, 2007; San Jose California USA. [doi: 10.1145/1240624.1240704]
28. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
29. Bierbooms JJ, van Haaren M, IJsselsteijn WA, de Kort YA, Feijt M, Bongers IM. Integration of online treatment into the "New normal" in mental health care in post-COVID-19 times: exploratory qualitative study. JMIR Form Res 2020 Oct 08;4(10):e21344 [FREE Full text] [doi: 10.2196/21344] [Medline: 33001835]
30. Korfage I. Digitale transformatie: techniekpaniek en onbewuste belemmeringen (Digital transformation: panic to technology and unconsious barriers). Erasmus Centrum voor Zorgbestuur (Erasmus Centre of Health Management). 2021. URL: https://www.erasmuscentrumzorgbestuur.nl/dynamic/media/24/documents/Inge%20Korfage-korfageinge_146472_43576783_Thesis.pdf [accessed 2001-09-21]
31. Spoelstra S. Van face to face naar zorg op afstand. Zorgvisie 2020 May 22;50(4):20-23 [FREE Full text] [doi: 10.1007/s41187-020-0367-x] [Medline: 32421051]
32. Greitzer FL, Kuchar OA, Huston K. Cognitive science implications for enhancing training effectiveness in a serious gaming context. J Educ Resour Comput 2007 Nov;7(3):2. [doi: 10.1145/1281320.1281322]
33. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Comput Entertain 2005 Jul;3(3):3. [doi: 10.1145/1077246.1077253]
34. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
35. Lindgren A, Chen F, Amdahl P, Chaikiat P. Using personas and scenarios as an interface design tool for advanced driver assistance systems. In: Universal Access in Human-Computer Interaction. Berlin, Heidelberg: Springer; 2007.
36. Human centred technology workshop 2003. University of Sussex. URL: http://www.sussex.ac.uk/informatics/cogslib/reports/csrp/csrp567.pdf#page=52 [accessed 2022-07-15]
37. Rondon S, Sassi FC, Furquim de Andrade CR. Computer game-based and traditional learning method: a comparison regarding students' knowledge retention. BMC Med Educ 2013 Mar 25;13:30 [FREE Full text] [doi: 10.1186/1472-6920-13-30] [Medline: 23442203]
38. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
39. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]

Abbreviations

eMH: e–mental health
GGzE: Stichting Geestelijke Gezondheidszorg Eindhoven en de Kempen

Edited by N Zary; submitted 04.11.21; peer-reviewed by A Pencer, N Gehring, A Joseph; comments to author 16.01.22; revised version received 24.03.22; accepted 13.05.22; published 27.07.22.

Please cite as:
Bierbooms J, Feijt MA, IJsselsteijn WA, Bongers IMB
Design of a Game-Based Training Environment to Enhance Mental Health Care Professionals' Skills in Using e–Mental Health: Multiple Methods User Requirements Analysis
JMIR Serious Games 2022;10(3):e34700
URL: https://games.jmir.org/2022/3/e34700
doi:10.2196/34700
PMID:35896032

©Joyce Bierbooms, Milou A Feijt, Wijnand A IJsselsteijn, Inge M B Bongers. Originally published in JMIR Serious Games (https://games.jmir.org), 27.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Biofeedback-Based Mobile App With Serious Games for Young Adults With Anxiety in the United Arab Emirates: Development and Usability Study

Mariam Almeqbaali[1], BSc; Sofia Ouhbi[1], PhD; Mohamed Adel Serhani[2], PhD; Leena Amiri[3], PhD; Reem K Jan[4], PhD; Nazar Zaki[1], PhD; Ayman Sharaf[1], BSc; Abdulla Al Helali[3], MBBS; Eisa Almheiri[3], MBBS

[1]Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

[2]Department of Information Systems and Security, College of Information Technology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

[3]Department of Psychiatry, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

[4]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

Corresponding Author:
Mohamed Adel Serhani, PhD
Department of Information Systems and Security
College of Information Technology
United Arab Emirates University
Sheik Khalifa Bin Zayed St
Al Ain, Abu Dhabi
United Arab Emirates
Phone: 971 3 7135572
Email: serhanim@uaeu.ac.ae

Abstract

Background: Following the outbreak of COVID-19, several studies have reported that young adults encountered a rise in anxiety symptoms, which could negatively affect their quality of life. Promising evidence suggests that mobile apps with biofeedback, serious games, breathing exercises, and positive messaging, among other features, are useful for anxiety self-management and treatment.

Objective: This study aimed to develop and evaluate the usability of a biofeedback-based app with serious games for young adults with anxiety in the United Arab Emirates (UAE).

Methods: This study consists of two phases: Phase I describes the design and development of the app, while Phase II presents the results of a usability evaluation by experts. To elicit the app's requirements during Phase I, we conducted (1) a survey to investigate preferences of young adults in the UAE for mobile games for stress relief; (2) an analysis of serious games for anxiety; and (3) interviews with mental health professionals and young adults in the UAE. In Phase II, five experts tested the usability of the developed app using a set of Nielsen's usability heuristics.

Results: A fully functional biofeedback-based app with serious games was co-designed with mental health professionals. The app included 4 games (ie, a biofeedback game, card game, arcade game, and memory game), 2 relaxation techniques (ie, a breathing exercise and yoga videos), and 2 additional features (ie, positive messaging and a mood tracking calendar). The results of Phase II showed that the developed app is efficient, simple, and easy to use. Overall, the app design scored an average of 4 out of 5.

Conclusions: The elicitation techniques used in Phase I resulted in the development of an easy-to-use app for the self-management of anxiety. Further research is required to determine the app's usability and effectiveness in the target population.

(JMIR Serious Games 2022;10(3):e36936) doi:10.2196/36936

XSL•FO
RenderX

KEYWORDS

connected mental health; mental health; anxiety; digital game; biofeedback; app; serious game; gaming; gamification; young adult; user-centered design; stress; stress relief; user-centred design; youth; user feedback; user experience; usability; user need; development; mHealth; mobile health

Introduction

Due to the COVID-19 pandemic and the preventive measures imposed to limit the spread of the virus, such as lockdowns and social distancing, young adults in the United Arab Emirates (UAE) have reported increased stress from work, home, and finances [1], which potentially puts them at risk for psychological issues including anxiety [2,3]. Before COVID-19, anxiety affected 20%-25% of adult patients in the UAE, who are considered a psychologically vulnerable group. The prevalence of anxiety grew following the emergence of COVID-19, and it now affects 55.7% of the adult population in the UAE [4]. Anxiety constitutes a normal response to stressful events; however, when it is persistent, excessive, and left untreated, it can lead to serious problems [5]. Anxiety has been recognized to negatively impact the quality of life and psychosocial functioning [6], and to be a predictor of a wide range of mental disorders [7]. Given the negative impact of anxiety on health, as well as the medical and financial burdens of mental health services, an effective method for managing anxiety in young adults is required.

One proposed solution is connected mental health, which is the use of information and communication technology in supporting mental health care. Due to the availability and popularity of mobile devices [8], mobile apps are one of the most popular connected mental health approaches, with many people reporting using their mobile devices to access health-related information and expressing an interest in using apps to track their anxiety [9,10]. A recent systematic review of highly rated anxiety apps has identified several apps that people with anxiety can use independently to practice relaxation and management methods [11]. Breathing exercises, yoga, and motivational quotes were among these management methods. Furthermore, half of these apps used gamification features to keep users engaged.

Studies have identified that games, especially commercial video games, have a significant impact on cognitive function, specifically on attentional control [12], and mental health including anxiety and stress [13]. In addition, games that are specifically designed to address a problem or teach a certain ability have had a lot of success [14]. Serious games are games with a main purpose other than pure entertainment, such as games used for education and health care purposes, with the added value of entertainment and competition. Researchers have demonstrated that serious games are comparable to traditional therapies, and for some users, can be more fun and acceptable [15]. Serious games also facilitate enhanced user engagement and increased motivation, which ultimately can improve treatment outcomes [16,17]. Consequently, serious games have been shown to improve cognitive functioning and aid in the treatment of a variety of mental disorders, including depression and posttraumatic stress disorder [18].

Biofeedback is another technique used for anxiety management, which focuses on helping people gain control over their physiological functions. Since anxiety is linked to physical symptoms, like increased heart rate and respiratory issues, biofeedback can be an effective tool for detecting and treating anxiety [19]. Incorporating biofeedback therapy into games may increase the benefits of both techniques in terms of anxiety reduction [20,21].

In general, mobile apps and serious games have shown promise in self-managing anxiety; however, few apps targeting anxiety are available for Arabic speakers [22,23]. Culturally adapting the app design in terms of language, culture, and context and aligning it with the standards and values of the target population can improve user acceptance toward mental health apps [24]. Therefore, the goal of this study was to develop and evaluate the usability of a biofeedback-based app with serious games for young adults with anxiety in the UAE. The design of the app is customized to employ elements from the UAE culture.

Methods

This study was conducted in two phases: (1) the development of a biofeedback-based app with serious games, and (2) the evaluation of the app's usability with experts.

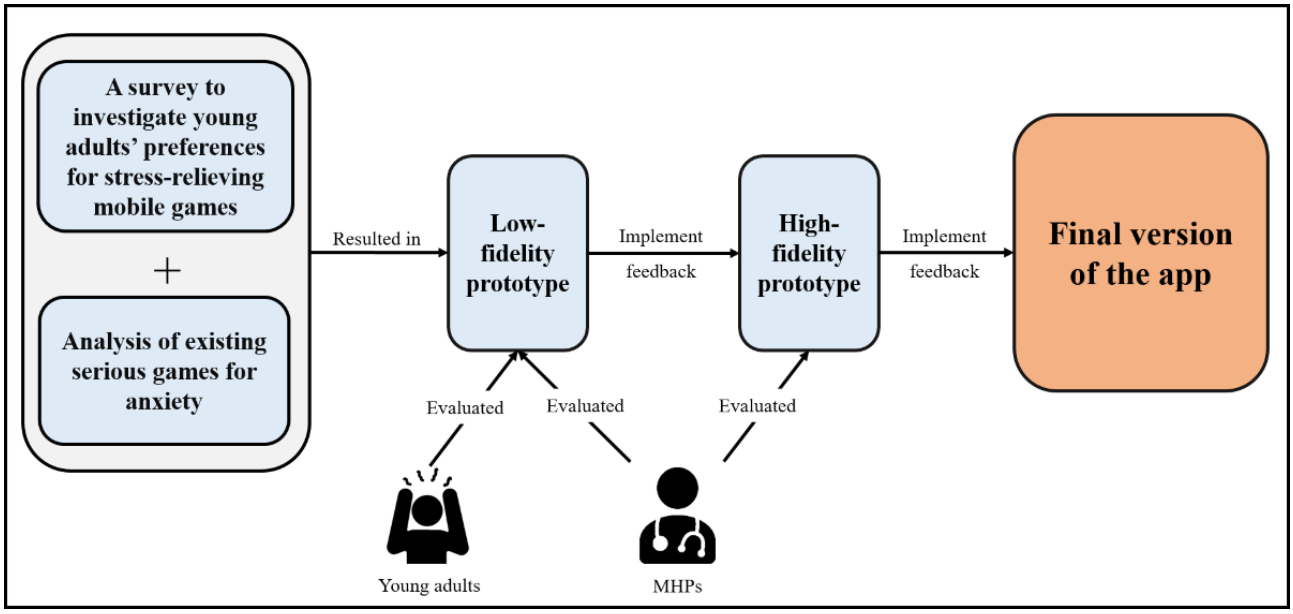
Phase I: Development of a Biofeedback-Based App With Serious Games

Overview

In this phase, we used a user-centered design approach, which focuses on understanding the perspective of the target users. Therefore, the app was developed by combining data collected from young adults in the UAE and the results from analyzing existing serious games for anxiety as well as feedback from mental health professionals (MHPs). Figure 1 presents the development process of our app, which consists of 3 main parts.

XSL•FO
RenderX

Figure 1. App development process. MHPs: mental health professionals.



Part 1: Investigation of Young Adults' Game Preferences for Stress Relief

The goal of this part was to analyze the preferences of young adults (18-37 years old) toward mobile games in terms of functionality and design for stress and anxiety relief. Therefore, an online questionnaire, prepared using Google Forms, was sent via social media and mailing lists to university students and participation was voluntary. The questionnaire was available online for a period of 2 weeks, from November 19, 2020, to December 3, 2020. The estimated time for completion was 5 minutes and the participants' answers were collected anonymously.

The online questionnaire consisted of 15 questions (7 multiple-choice questions, 3 yes/no questions, and 5 open questions). Four of the multiple-choice questions were multi-select. The questions were divided into three sections: demographics, stress-related, and game-related questions. For the student community, we used "stress" to replace "anxiety," as participants might be unaware of the real meaning of anxiety and might refer to their anxiety as stress. Furthermore, high levels of stress can predict anxiety [25].

Part 2: Analysis of Existing Serious Games for Anxiety

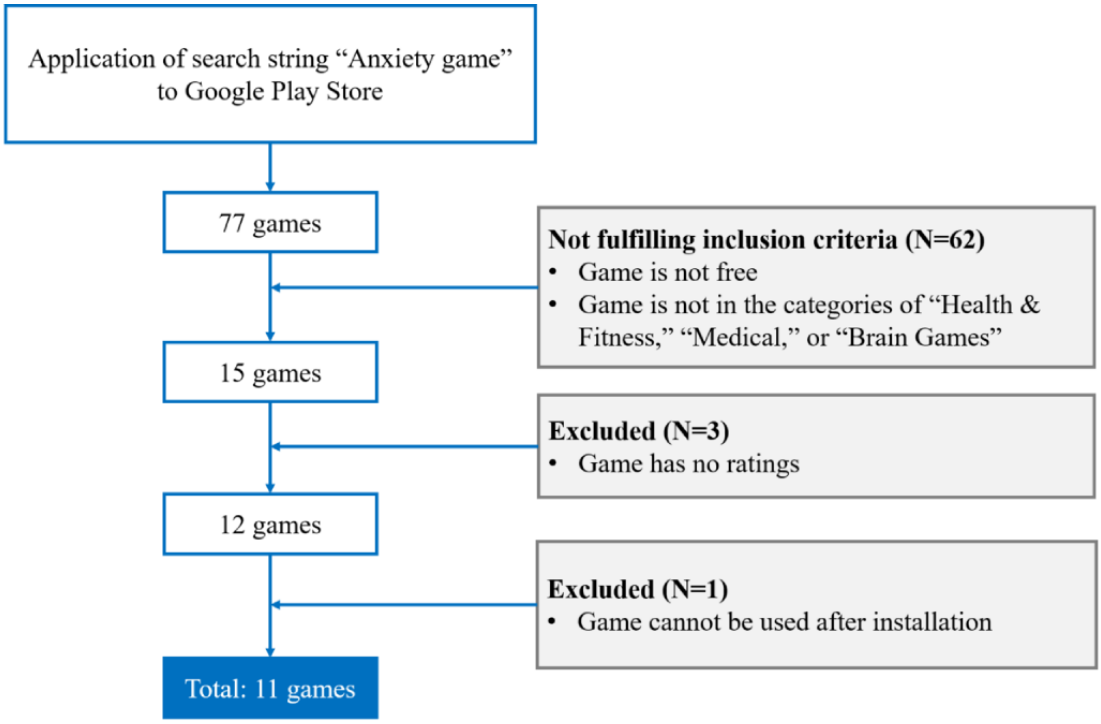
The goal of this analysis was to observe two major issues: the characteristics of available serious games on mobile phones, and their overall design. Therefore, a systematic review was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [26]. A general search query composed of the terms "anxiety" and "game" was used and was automatically applied to the titles and descriptions of the games available in Google Play during March 2021. Each game from the search results was reviewed before deciding whether it should be included or excluded from the final selection. The following inclusion criteria were applied: (1) anxiety-related games in the Google Play store, (2) games that have a free version, and (3) games within the categories "Health & Fitness," "Medical," or "Brain Games."

The categories were intended to highlight serious games that were primarily concerned with serious topics such as health and well-being, rather than games that were pure entertainment. Brain games, known as cognitive training, are specialized activities that mainly focus on enhancing attention and working memory [27], in turn having a role in relieving anxiety. To identify the final selection that would be examined, the following exclusion criteria were applied to the candidate games: (1) games that do not have ratings, and (2) games that could not be used after installation.

A total of 11 Android games were included in the final selection after the application of the aforementioned criteria (Figure 2). Each selected game was installed and examined on a Galaxy Note 9 (Android 10). A data extraction form was created in Microsoft Excel (Microsoft Corp). The games were classified into categories including puzzle, adventure, arcade, and simulation.

XSL•FO
RenderX

Figure 2. Selection process.



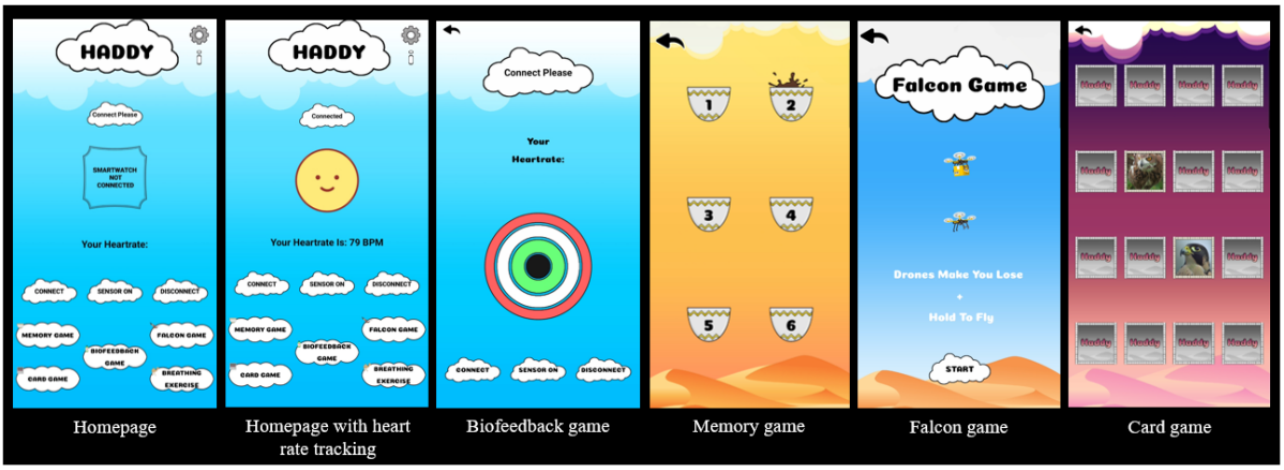
Part 3: Co-design With Mental Health Professionals

Two MHPs from the UAE were invited to participate in the co-design of the app. Both MHPs work as teaching assistants at UAEU College of Medicine, and are currently psychiatry and psychotherapy residents in Germany. They were chosen because they had previously worked with patients with anxiety in the UAE. Two evaluation cycles were conducted to refine the low-fidelity and high-fidelity prototypes. The evaluation was in the form of semistructured interviews with MHPs. The interviews were conducted online via Microsoft Teams and were digitally recorded to facilitate data analysis.

During the first cycle, the meeting started with a quick review of the project proposal, followed by a presentation of the low-fidelity prototype. Figure 3 presents pictures of the low-fidelity prototype screens presented to MHPs. The low-fidelity prototype was created based on the findings of Part 1 (Multimedia Appendix 1) and Part 2 (Multimedia Appendix 2) of Phase I. The main game, the biofeedback game, was developed in response to Part 1 findings indicating that young adults' willingness to play games during stressful conditions increased if the game tracked heart rate and assisted with breathing exercises. Three key questions concerning the advantages, disadvantages, and potential modifications were asked after each proposed feature of the app. The findings were analyzed after the meeting to modify the prototype before cycle 2 with MHPs. Similarly, the low-fidelity prototype was also assessed by 6 university students aged between 18 and 25 years, who self-identified as having study anxiety. The interview was conducted and feedback on the prototype was solicited from the students. The prototype was well received by the students, with praise for the animation, types of games, and ease of use.

XSL•FO
RenderX

Figure 3. Screenshots of the low-fidelity prototype.



During the second cycle, MHPs evaluated the high-fidelity prototype, which was developed using Android Studio, and Tizen Studio for the smartwatch app. The Android Application Package (APK) for the app was emailed to the MHPs involved in cycle 1, along with a video that demonstrated the app's primary features, before the second meeting. During the interview, 5 guiding questions were asked to cover topics including efficacy, preferences, design and graphics, engagement, and additional modifications. The following quotes demonstrate some of the feedback we received from MHPs after the second cycle:

- "Adding sound effects to the breathing exercise would have a positive impact on allowing the clients to relax"
- "The games would help with concentration and being present in the here and now. This will improve the overall mental state."
- "The app will be useful for people with moderate anxiety"
- "Clients with severe anxiety would not be in a state of wanting to do such exercises"

Phase II: Usability Evaluation

Data acquired from Phase I in this study resulted in an app named *Haddy*, which means "calm down" or "relax" in the Arabic Emirati dialect. In this phase, experts were involved in evaluating the usability of the final version of the app. In order to participate in this evaluation, an expert needed to have at least 3 years of experience in usability testing and software engineering. The research team used personal contacts to recruit 5 experts for this study. For the heuristic evaluation, 5 participants were considered a sufficient sample size [28].

Experts were required to download the app from the Google Play Store to evaluate the app. The evaluation was conducted using a questionnaire with a total of 30 items, 19 of which were based on Nielsen's usability heuristics. Match between system and the real world, visibility of system status, memory, minimalist design, error prevention, consistency, user control, and flexibility are among the heuristics used in this evaluation. A 5-point Likert scale was used for the assessment (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree). In addition, 9 items were included to score each feature of the app on a scale from 1 to 5, and 2 open questions were used to highlight the app's advantages and problems related to the usability of the interface design.

Ethics Approval

Ethical approval for the study was obtained from the relevant authorities at the United Arab Emirates University (ERS_2020_6156).

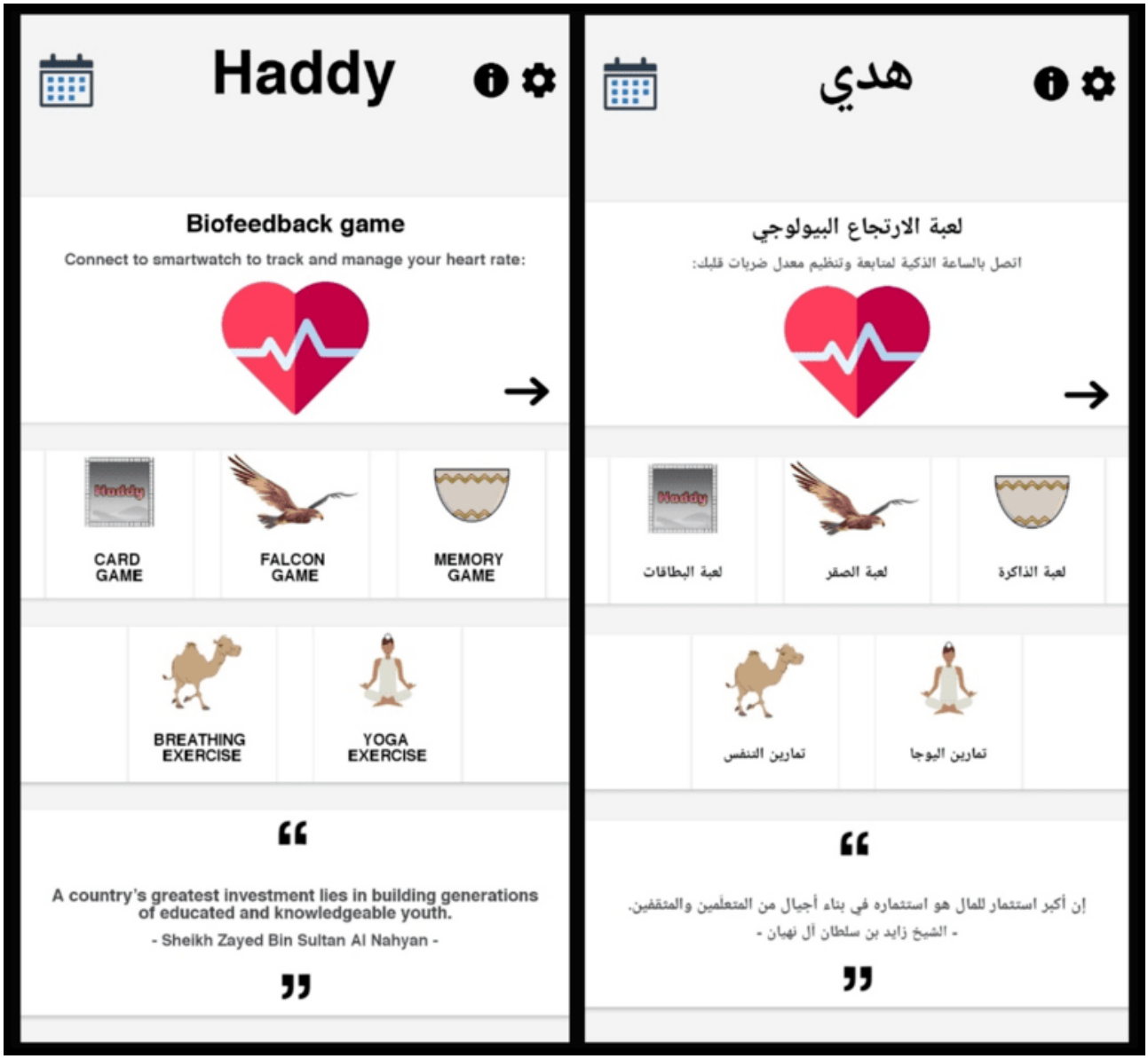
Results

Phase I: Development of a Biofeedback-Based Mobile App With Serious Games

The final version of the app (Figure 4) has the following key features: (1) biofeedback-based game, (2) card game, (3) falcon game, (4) memory game, (5) breathing exercise, (6) yoga videos, (7) positive messaging, and (8) mood tracking calendar. Since it is directed at people in the UAE, the app uses both Arabic and English languages.

XSL•FO
RenderX

Figure 4. The final version of the app's homepage in English and Arabic.



The biofeedback game (Figure 5) was designed to help the player gain control over physical symptoms of anxiety. It requires the use of a smartwatch for heart rate tracking. To start the game, the player must first connect to the smartwatch by tapping the connect button. Prior to that, the user must put on the smartwatch and pair it with the phone via Bluetooth. After a few seconds of connection, the heart rate should appear. Heart rate is represented by colored circles: green (normal, at rest), red (high, anxious), and blue (low, dangerous). The normal heart rate falls between 60 and 100 beats per minute according to MHPs. When the color is not green, the player must relax by practicing activities such as breathing until the color goes back to green.

The card game (Figure 6A) is a classic mini-game with the goal of turning 2 matching cards at the same time. This game was designed to divert the user's attention away from anxious thoughts. The game consists of three difficulty levels: easy (4×4 grid), medium (5×4 grid), and difficult (6×4 grid). The graphics on the cards portray animal species from the UAE, with a desert in the background. The game has no time or movement restrictions and it ends when all of the cards are acquired.

XSL•FO
RenderX

Figure 5. User interfaces for the biofeedback game.

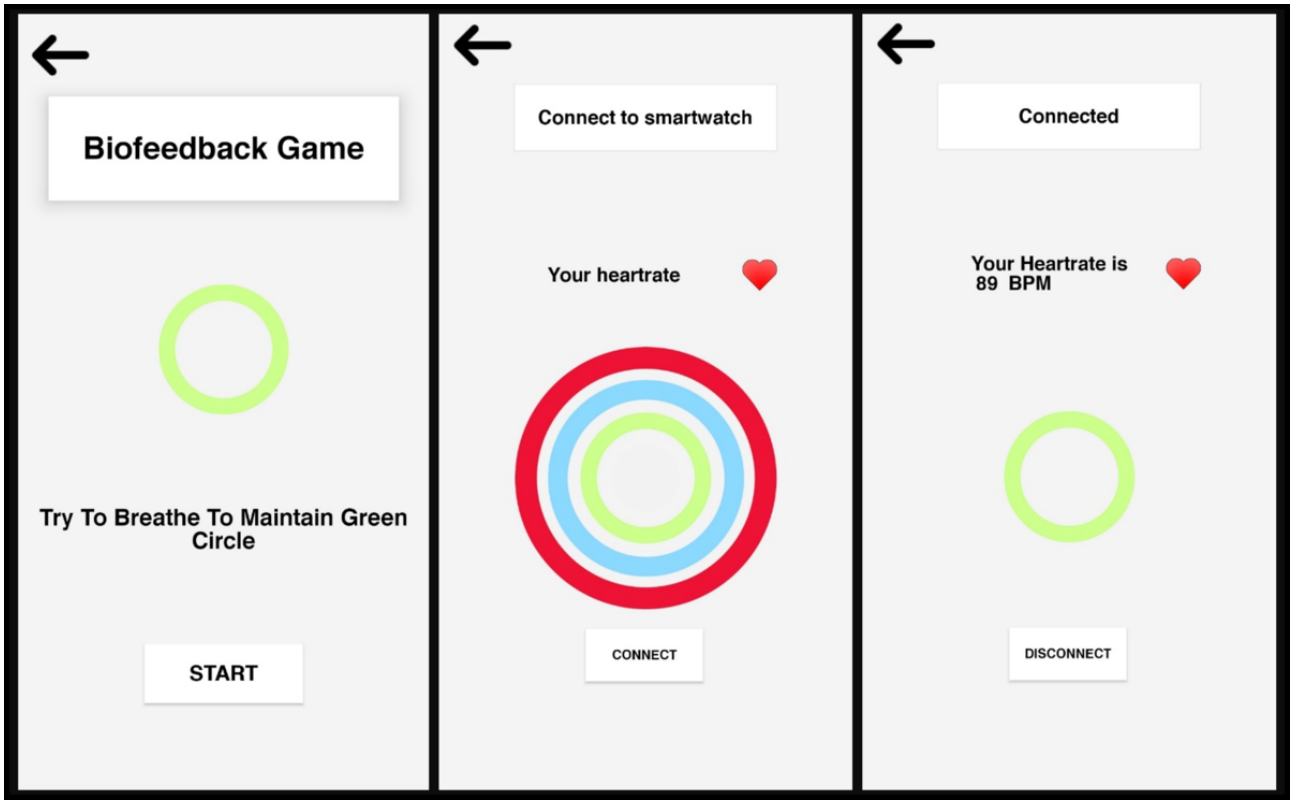
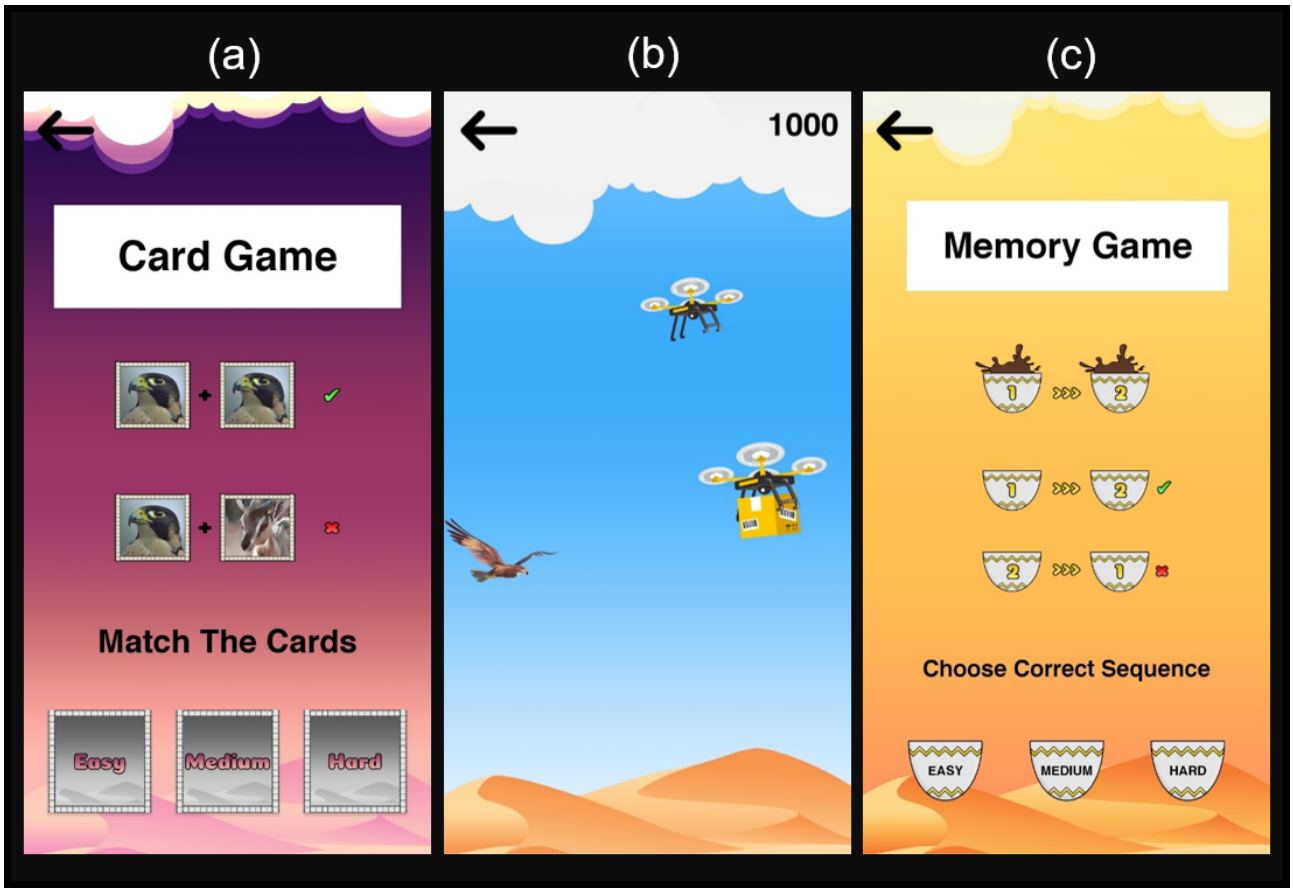


Figure 6. Mini-games including (A) the card game, (B) the falcon game, and (C) the memory game.



The falcon game (Figure 6B) is a side-view perspective arcade game used to divert someone's attention away from negative thoughts and toward adjusting the falcon's position. In this game, the falcon must fly through deadly drones to win. To keep the falcon flying, the player must keep pressing on the screen. Every 10 seconds, the score increases. If the falcon hits

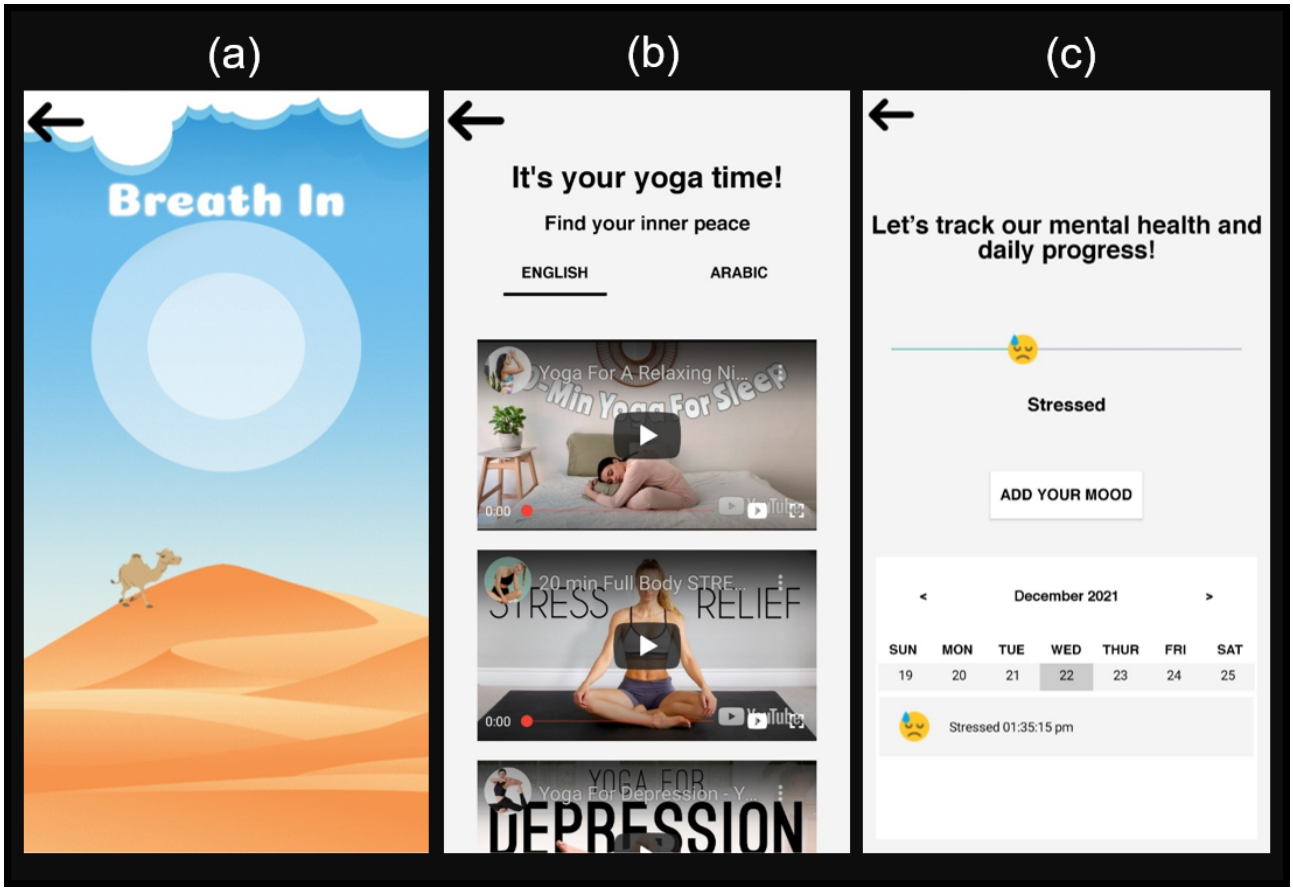
XSL•FO
RenderX

a drone, the game will end. If it does not, the game will continue, with the falcon's speed increasing as the game proceeds. The highest score of the player will be saved.

The memory game (Figure 6C) is another game offered by the app, with a similar principle to the card game. A set of Arabic coffee cups is presented, and a sequence of pouring coffee into each cup begins one by one. The player must memorize and match the sequence in order to win the game. If the player chooses the incorrect order, the game will end. This game also presents three difficulty levels: easy (4-item sequence), medium (6-item sequence), and difficult (8-item sequence).

The breathing exercise (Figure 7A) is a guided breathing exercise that instructs the user on when to inhale and exhale. Users are able to customize the duration of the exercise. They have four options: 2 minutes (which is the shortest), 5 minutes, 10 minutes, and 15 minutes (the longest). We added relaxing music to the breathing exercise based on feedback from the MHPs.

Figure 7. Other app functionalities include (A) a breathing exercise, (B) yoga videos, and (C) a mood tracking calendar.



Yoga videos (Figure 7B) were included in both English and Arabic. There are a total of 12 yoga videos available, half in Arabic and half in English. The videos were obtained from YouTube channels; hence, they can only be viewed with internet access.

Positive messaging or quotes (Figure 4) were placed at the bottom of the homepage screen. Some of the quotes come from well-known Emirati public figures, such as the UAE's beloved first president, Sheikh Zayed bin Sultan. When the user reopens the app, a new quote will appear.

A mood tracking calendar (Figure 7C) is a simple feature for people who wish to keep track of their mood on a daily basis. This can be useful when reflecting back on how one has been feeling over a period of time. Different moods can be observed if the user slides the progress bar. Users can define their current mood by selecting 1 of the 9 moods (tired, lonely, bored, stressed, I don't know, calm, happy, excited, and wonderful) indicated with an icon and text. The user can add different moods in a single day by clicking the "Add mood" button. The moods will appear in chronological order in a weekly calendar.

Phase II: Usability Evaluation

All 5 experts completed the usability evaluation questionnaire for the final version of *Haddy*. Table 1 shows the mean usability score from the experts' point of view, with the consistency of the app scoring the highest. On the other hand, the ability to hide or display information, which is an aspect of minimalist design, was clearly low. In addition, both items of flexibility, including customization of functionalities and properties of interfaces, scored less than 4.

XSL•FO
RenderX

Table 1. Usability heuristics score (n=5).

Question	Mean (SD)[a]
Match between system and the real world	
1. The app's terminology is exposed in a familiar way to the user	4.6 (0.55)
2. The app icons and images are exposed in a familiar way to the user	4.4 (0.55)
3. The app's screens are presented in the most logical way	4.0 (1.00)
Visibility of system status	
4. There is visual feedback that highlights the options that can be selected	3.6 (1.14)
5. There is no observable delay in the response time of the system	4.8 (0.45)
6. The app provides feedback to the user about its current status	4.4 (0.55)
Recognition rather than recall	
7. There is enough information displayed at each step in the tasks	4.4 (0.55)
8. The user can easily navigate back and forth without the need to remember each step	4.6 (0.55)
Esthetic and minimalist design	
9. The duplicated information is eliminated	4.8 (0.48)
10. It is possible to display or hide information (ie, expanding, collapsing lists)	2.4 (1.52)
11. The app design is minimalistic	4.4 (0.55)
Error prevention	
12. The app's design minimizes the possibility of errors	4.6 (0.55)
13. The app warns the user of the error type in plain language	3.4 (1.52)
Consistency and standards	
14. The app remains consistent with elements that perform the same actions	4.8 (0.45)
15. The names of the options in the homepage are consistent in grammatical style and terminology	4.8 (0.45)
User control and freedom	
16. It is always possible to cancel the actioning of a task	4.2 (1.30)
17. It is easy for the user to undo and redo the actions	3.8 (1.30)
Flexibility and efficiency of use	
18. The user can change some properties of the interface	3.6 (0.89)
19. The app's functionality can be customized by the user	3.2 (1.48)
Rating scale of all items (total=95)	78.8 (9.58)

[a]All items were rated on a 5-point scale from 1=strongly disagree to 5=strongly agree.

According to the experts' evaluation, the positive messaging and the card game were the most well-designed features of the app. Biofeedback, the falcon game, and the yoga videos, in contrast, had the lowest scores of all the features, but their scores remained over 4. Overall, the app design scored 4 out of 5. Table 2 shows the rest of the findings.

XSL•FO
RenderX

Table 2. Rating score for each feature in the app.[a]

Question	Participants					Mean (SD)
	P1	P2	P3	P4	P5	
Biofeedback game	4	4	3.5	5	4	4.1 (0.55)
Card game	4	5	4.5	4	5	4.5 (0.50)
Falcon game	4.5	5	4	4	3	4.1 (0.74)
Memory game	4.5	5	4	4	4	4.3 (0.45)
Breathing exercise	5	3	4.5	5	4	4.3 (0.84)
Yoga videos	5	4	3.5	4	4	4.1 (0.55)
Positive messaging	5	5	3.5	5	5	4.7 (0.67)
Calendar mental status tracking	5	3	3.5	5	5	4.3 (0.97)
Overall app design	4.5	4	4	4.5	3	4.0 (0.61)

[a]All items were rated out of 5.

The last two questions in the usability questionnaire were designed to learn what experts liked and did not like about the app design. All the experts agreed that the app is simple and easy to use. One expert pointed out that the design suits the Arab region. On the other hand, experts identified some disadvantages, such as the amount of text displayed at once, improper error handling, and inconsistency with the display of icons. Along with the feedback, suggestions on how to improve the current version of the app were acquired and presented in Table 3.

Table 3. Suggestions provided by experts for improvement of the current version of the app.

Category	Suggestions
Minimization	• Reduce the text on the instructions page of the breathing exercise, as the page is crowded.
Error handling	• Display an error message to warn the user when the smartwatch is not detected in the biofeedback game.
Design	• Improve the button design to make it clear where to click.
Consistency	• Present the icon of the 'mood tracking calendar' in the same way as other features are displayed on the homepage. • Improve the position of icons on the homepage.
Customization	• Add "delete" and "edit" options in the mood tracking calendar. • Allow for customization of the time duration in the breathing exercise. • Provide mute/unmute options for the music in the breathing exercise. • Allow full-screen viewing of yoga videos.

Discussion

Principal Findings

The goal of this study was to develop and test the usability of a biofeedback-based app with serious games for anxiety management for young adults in the UAE. According to the results of the questionnaire, games with heart rate tracking and breathing exercises have the potential to increase young adults' interest in gaming as a stress-relieving method. This app was co-designed with MHPs in order to ensure the content's integrity. The involvement of MHPs in the design process is thought to improve the efficacy of apps and to boost users' trust in apps. A study has shown that the number of installs of anxiety apps that included MHPs was significantly higher than the number of installs of anxiety apps that did not include MHPs [29]. The target population for this app was young adults with mild to moderate anxiety symptoms, and not those with severe anxiety, since according to the MHPs involved in this study, those with severe anxiety were less likely to use this method of "self-help." This is supported by findings by Christensen et al [30] that people with high anxiety symptoms were more likely to drop out of internet-based therapies. MHPs considered games as an effective technique to boost user concentration and shift attention away from anxious thoughts, resulting in better anxiety reduction outcomes. Additionally, audio was incorporated into the breathing exercise because sensory experiences such as music were emphasized for relaxation and reduction of anxiety [31,32].

Good usability ratings and positive feedback from experts may be in part due to the elicitation techniques used in Phase I, which included obtaining information from potential users, existing solutions, and MHPs. Experts recommended increasing customization, both for the interface and for particular features. Many users of anxiety and other mental health apps place more emphasis on customization [33]. Experts were generally satisfied with the app and found it easy to use, with only minor flaws

XSL•FO
RenderX

that will be addressed in the next version. The findings of this study resulted in a future modification to facilitate the process of testing the app in young adults with anxiety.

Recommendations

We believe that our study will provide developers and practitioners insight into potential and acceptable content and features, as well as different considerations while developing an app for young adults with anxiety in the UAE. Based on the findings, we recommend app developers to consider (1) the involvement of users and MHPs in the requirement elicitation process to improve usability; (2) reducing the amount of information presented in the user interface; (3) customization of the user interface as well as some elements of proposed features; and (4) handling all errors that may arise and ensuring the user is aware of them.

Limitations

This study has a few limitations. The first limitation is that the usability of the high-fidelity prototype was evaluated by experts rather than young adults with anxiety; however, these results remain important, as the experts' feedback will ensure that the app is suitable for use with the target population. Another limitation is that this study focused on the app's interface conviviality, not the efficacy of the app in reducing anxiety symptoms, as this would require participant recruitment. We hope to overcome these limitations in future work by conducting a usability and efficacy evaluation with young adults with anxiety in the UAE. Finally, this app was initially developed for Android, but we intend to expand it to iOS in the future.

Conclusions

This study has resulted in the development of a culturally sensitive biofeedback-based app with serious games to help young adults with anxiety in the UAE. *Haddy* was developed in collaboration with MHPs and tested by experts. The app has been well received by MHPs, with positive feedback about its ability to help with anxiety management. Experts who have evaluated the use of the app reported that it is simple and easy to use. Experts' feedback will be included in the next version of the app. We intend to conduct a study to investigate the usability and impact of the future version of the app on a sample of young adults with anxiety in the UAE.

Acknowledgments

This work is part of the Abu Dhabi Young Investigator Award (AYIA) 2019 (#AYIA19-001) awarded by the Abu Dhabi Research and Development Authority (ADRDA) to SO. We would like to thank all the students who contributed to this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions and answers to the questionnaire that investigates young adults' preferences regarding games for stress relief.
[DOCX File , 14 KB - games_v10i3e36936_app1.docx]

Multimedia Appendix 2
Results of the analysis of Android mobile games for anxiety.
[DOCX File , 13 KB - games_v10i3e36936_app2.docx]

References

1. Cheikh Ismail L, Mohamad MN, Bataineh MF, Ajab A, Al-Marzouqi AM, Jarrar AH, et al. Impact of the coronavirus pandemic (COVID-19) lockdown on mental health and well-being in the United Arab Emirates. Front Psychiatry 2021;12:633230 [FREE Full text] [doi: 10.3389/fpsyt.2021.633230] [Medline: 33796033]
2. Drissi N, Alhmoudi A, Al Nuaimi H, Alkhyeli M, Alsalami S, Ouhbi S. Investigating the impact of COVID-19 lockdown on the psychological health of university students and their attitudes toward mobile mental health solutions: Two-part questionnaire study. JMIR Form Res 2020 Oct 20;4(10):e19876 [FREE Full text] [doi: 10.2196/19876] [Medline: 32969340]
3. Saddik B, Hussein A, Sharif-Askari FS, Kheder W, Temsah MH, Koutaich RA, et al. Increased levels of anxiety among medical and non-medical university students during the COVID-19 pandemic in the United Arab Emirates. Risk Manag Healthc Policy 2020;13:2395-2406 [FREE Full text] [doi: 10.2147/RMHP.S273333] [Medline: 33177898]
4. Thomas J, Barbato M, Verlinden M, Gaspar C, Moussa M, Ghorayeb J, et al. Psychosocial correlates of depression and anxiety in the United Arab Emirates during the COVID-19 pandemic. Front Psychiatry 2020;11:564172 [FREE Full text] [doi: 10.3389/fpsyt.2020.564172] [Medline: 33240122]
5. Anxiety Disorders. WebMD. URL: https://www.webmd.com/anxiety-panic/guide/anxiety-disorders [accessed 2021-11-11]
6. Bourland SL, Stanley MA, Snyder AG, Novy DM, Beck JG, Averill PM, et al. Quality of life in older adults with generalized anxiety disorder. Aging & Mental Health 2000 Nov;4(4):315-323. [doi: 10.1080/713649960]
7. Essau CA, Lewinsohn PM, Olaya B, Seeley JR. Anxiety disorders in adolescents and psychosocial outcomes at age 30. J Affect Disord 2014 Jul;163:125-132 [FREE Full text] [doi: 10.1016/j.jad.2013.12.033] [Medline: 24456837]

XSL•FO
RenderX

8. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: Systematic review. J Med Internet Res 2020 Jan 06;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
9. Atallah N, Khalifa M, El Metwally A, Househ M. The prevalence and usage of mobile health applications among mental health patients in Saudi Arabia. Comput Methods Programs Biomed 2018 Mar;156:163-168. [doi: 10.1016/j.cmpb.2017.12.002] [Medline: 29428068]
10. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
11. Drissi N, Ouhbi S, Janati Idrissi MA, Ghogho M. An analysis on self-management and treatment-related functionality and characteristics of highly rated anxiety apps. Int J Med Inform 2020 Sep;141:104243 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104243] [Medline: 32768994]
12. Kowal M, Conroy E, Ramsbottom N, Smithies T, Toth A, Campbell M. Gaming Your Mental Health: A Narrative Review on Mitigating Symptoms of Depression and Anxiety Using Commercial Video Games. JMIR Serious Games 2021 Jun 16;9(2):e26575 [FREE Full text] [doi: 10.2196/26575] [Medline: 34132648]
13. Barr M, Copeland-Stewart A. Playing Video Games During the COVID-19 Pandemic and Effects on Players' Well-Being. Games and Culture 2021 May 06;17(1):122-139. [doi: 10.1177/15554120211017036]
14. Griffiths MD. The therapeutic and health benefits of playing video games. In: The Oxford Handbook of Cyberpsychology. Oxford: Oxford University Press; 2019.
15. Horne-Moyer HL, Moyer BH, Messer DC, Messer ES. The use of electronic games in therapy: a review with clinical implications. Curr Psychiatry Rep 2014 Dec;16(12):520 [FREE Full text] [doi: 10.1007/s11920-014-0520-6] [Medline: 25312026]
16. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 01;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
17. Boendermaker WJ, Gladwin TE, Peeters M, Prins PJM, Wiers RW. Training working memory in adolescents using serious game elements: Pilot randomized controlled trial. JMIR Serious Games 2018 May 23;6(2):e10 [FREE Full text] [doi: 10.2196/games.8364] [Medline: 29792294]
18. Lau HM, Smit JH, Fleming TM, Riper H. Serious Games for Mental Health: Are They Accessible, Feasible, and Effective? A Systematic Review and Meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
19. Alneyadi M, Drissi N, Almeqbaali M, Ouhbi S. Biofeedback-based connected mental health interventions for anxiety: systematic literature review. JMIR mHealth uHealth 2021 Apr 22;9(4):e26038 [FREE Full text] [doi: 10.2196/26038] [Medline: 33792548]
20. Knox M, Lentini J, Cummings T, McGrady A, Whearty K, Sancrant L. Game-based biofeedback for paediatric anxiety and depression. Ment Health Fam Med 2011 Sep;8(3):195-203 [FREE Full text] [Medline: 22942901]
21. Schuurmans AAT, Nijhof KS, Engels RCME, Granic I. Using a videogame intervention to reduce anxiety and externalizing problems among youths in residential care: An initial randomized controlled trial. J Psychopathol Behav Assess 2018;40(2):344-354 [FREE Full text] [doi: 10.1007/s10862-017-9638-2] [Medline: 29937625]
22. Alhuwail D, Albaj R, Ahmad F, Aldakheel K. The state of mental digi-therapeutics: A systematic assessment of depression and anxiety apps available for Arabic speakers. Int J Med Inform 2020 Mar;135:104056. [doi: 10.1016/j.ijmedinf.2019.104056] [Medline: 31918341]
23. Ahmed A, Ali N, Aziz S, Abd-alrazaq AA, Hassan A, Khalifa M, et al. A review of mobile chatbot apps for anxiety and depression and their self-care features. Computer Methods and Programs in Biomedicine Update 2021;1:100012. [doi: 10.1016/j.cmpbup.2021.100012]
24. Alsswey AH, Al-Samarraie H, El-Qirem FA, Alzahrani AI, Alfarraj O. Culture in the design of mHealth UI. EL 2020 Mar 27;38(2):257-272. [doi: 10.1108/el-04-2019-0097]
25. Stress in America: Generation Z. American Psychological Association. 2018. URL: https://www.apa.org/news/press/releases/stress/2018/stress-gen-z.pdf [accessed 2021-11-11]
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
27. Al-Thaqib A, Al-Sultan F, Al-Zahrani A, Al-Kahtani F, Al-Regaiey K, Iqbal M, et al. Brain training games enhance cognitive function in healthy subjects. Med Sci Monit Basic Res 2018 Apr 20;24:63-69 [FREE Full text] [doi: 10.12659/msmbr.909022] [Medline: 29674605]
28. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam p. 206-213. [doi: 10.1145/169059.169166]
29. Drissi N, Ouhbi S. On the involvement of mental healthcare professionals in the co-design of highly-rated anxiety apps. HEALTHINF 2021:281-287. [doi: 10.5220/0010183102810287]

XSL•FO
RenderX

30. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
31. Ventura T, Gomes MC, Carreira T. Cortisol and anxiety response to a relaxing intervention on pregnant women awaiting amniocentesis. Psychoneuroendocrinology 2012 Jan;37(1):148-156. [doi: 10.1016/j.psyneuen.2011.05.016] [Medline: 21705148]
32. Voss JA, Good M, Yates B, Baun MM, Thompson A, Hertzog M. Sedative music reduces anxiety and pain during chair rest after open-heart surgery. Pain 2004 Nov;112(1-2):197-203. [doi: 10.1016/j.pain.2004.08.020] [Medline: 15494201]
33. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]

Abbreviations

MHP: mental health professional
UAE: United Arab Emirates

Edited by N Zary; submitted 31.01.22; peer-reviewed by A Hashim, N Chalghaf; comments to author 25.03.22; revised version received 06.04.22; accepted 12.06.22; published 02.08.22.

Please cite as:
Almeqbaali M, Ouhbi S, Serhani MA, Amiri L, Jan RK, Zaki N, Sharaf A, Al Helali A, Almheiri E
A Biofeedback-Based Mobile App With Serious Games for Young Adults With Anxiety in the United Arab Emirates: Development and Usability Study
JMIR Serious Games 2022;10(3):e36936
URL: https://games.jmir.org/2022/3/e36936
doi:10.2196/36936
PMID:35916692

©Mariam Almeqbaali, Sofia Ouhbi, Mohamed Adel Serhani, Leena Amiri, Reem K Jan, Nazar Zaki, Ayman Sharaf, Abdulla Al Helali, Eisa Almheiri. Originally published in JMIR Serious Games (https://games.jmir.org), 02.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Biomarkers for Well-being Through Exergame Interactions: Exploratory Study

Despoina Petsani[1], MSc; Evdokimos Konstantinidis[1], PhD; Aikaterini-Marina Katsouli[1], MEd; Vasiliki Zilidou[1], PhD; Sofia B Dias[2], PhD; Leontios Hadjileontiadis[3,4], PhD; Panagiotis Bamidis[1], PhD

[1]Medical Physics and Digital Innovation Laboratory, Faculty of Health Sciences, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]Centro Interdisciplinar de Estudo da Performance Humana, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal

[3]Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

[4]Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:
Panagiotis Bamidis, PhD
Medical Physics and Digital Innovation Laboratory
Faculty of Health Sciences, School of Medicine
Aristotle University of Thessaloniki
Medical Physics and Digital Innovation Laboratory
Thessaloniki, 54124
Greece
Phone: 30 2310999310
Email: bamidis@med.auth.gr

Abstract

Background: Ecologically valid evaluations of patient states or well-being by means of new technologies is a key issue in contemporary research in health and well-being of the aging population. The in-game metrics generated from the interaction of users with serious games (SG) can potentially be used to predict or characterize a user's state of health and well-being. There is currently an increasing body of research that investigates the use of measures of interaction with games as digital biomarkers for health and well-being.

Objective: The aim of this paper is to predict well-being digital biomarkers from data collected during interactions with SG, using the values of standard clinical assessment tests as ground truth.

Methods: The data set was gathered during the interaction with patients with Parkinson disease with the webFitForAll exergame platform, an SG engine designed to promote physical activity among older adults, patients, and vulnerable populations. The collected data, referred to as in-game metrics, represent the body movements captured by a 3D sensor camera and translated into game analytics. Standard clinical tests gathered before and after the long-term interaction with exergames (preintervention test vs postintervention test) were used to provide user baselines.

Results: Our results showed that in-game metrics can effectively categorize participants into groups of different cognitive and physical states. Different in-game metrics have higher descriptive values for specific tests and can be used to predict the value range for these tests.

Conclusions: Our results provide encouraging evidence for the value of in-game metrics as digital biomarkers and can boost the analysis of improving in-game metrics to obtain more detailed results.

(JMIR Serious Games 2022;10(3):e34768) doi:10.2196/34768

KEYWORDS

serious games; machine learning; physical well-being; cognitive well-being

XSL•FO
RenderX

Introduction

Background and Rationale

Serious games (SG) for health are games that aim to provide additional value for players other than mere entertainment and specifically deal with aspects of physical, mental, and social well-being [1], following the World Health Organization's definition for health [2]. Most findings suggest that SG for health are effective interventions for increasing older people's mental and physical health and well-being, but there are strong variations in the outcomes and measures used to demonstrate impact [3,4] or assess users during gameplay.

A regularly used method for evaluating the impact of SG interventions is the use of external questionnaires for each participant, one before playing (before the test) and another after going through a series of playing sessions (after the test). This methodology is widely accepted in the health and well-being domain to evaluate the effectiveness of SG [5] but fails to provide more detailed and reliable information [6] on user's state of health and well-being. Furthermore, the collection of ex situ data such as before the test and after the test requires human resources and is opposed to a more natural experience [7], as they are usually collected in laboratory or clinical settings by clinical experts. Questionnaires, interviews, or test batteries are often used as assessment tools, and when administered, can cause stress to the interviewee, threatening accuracy and ecological validity [8]. On the contrary, the fact that SG can be administered in any setting and in an enjoyable way accounts for an ecologically valid environment where diagnostic processes could become unobtrusive [9]. Thus, the efficiency of using SG for evaluating the state of health and well-being of players should be further investigated, as it is one of the less studied subjects in SG research [10].

Digital SG enable in situ data capturing, which can reveal new insights, except from the usual retrospective analysis of the intervention results. The term *in-game metrics* is usually used to describe in situ data that are collected during the interaction of the user with the SG. In-game metrics can range from the time required to perform a task in the game to a complicated calculation of a score or lower-level data monitoring user interaction. In-game metrics can be very diverse, as there are currently no standards or guidelines of what data should be collected and for what purpose [11]. In-game metrics can be used to adapt to difficulty [12,13], monitor a user's behavior [14], and evaluate learning progress [15]. Understanding and exploiting in-game metrics is challenging, and the creation of new methods for interpreting in-game metrics into meaningful insights can strengthen SG research and effectiveness [16,17].

The in-game metrics that are generated can be a rich source of information and insights that can potentially be used to predict or characterize a user's state of health and well-being. The study by Regan et al [18] supports that the data generated during the gameplay, namely the in-game metrics, are promising digital biomarkers for mental health, whereas the study by Staiano and Calvert [19] also argues for the potential of SG to assess physical health. The identification of deviations from the norm in a gameplay or the correlation of in-game metrics with ex situ clinical data can be indicators of mental or physical decline. The reliability and validity of in-game metrics should be further investigated to specify at what extent they can capture changes in health and well-being within research studies. This proof will strengthen the value of in-game metrics as efficient, unobtrusive, and comprehensive research tools to measure participants' health.

The current work is part of an extended, holistic body of work, the Long Lasting Memories Care (NCT02313935), concerning SG for the physical and cognitive improvement of older adults and other vulnerable populations. The exergame platform of Long Lasting Memories Care has proven to significantly improve strength, flexibility, endurance, and balance in older adults [20]. Furthermore, a previous study conducted by Konstantinidis et al [21] demonstrated the value of in-game metrics generated from body movement interactions with the exergame for detecting cognitive decline, and the study by Anagnostopoulou et al [22] demonstrated evidence of improving the functional architecture of the brain in adults with Down syndrome.

Related Work

SG have already been used for modeling or characterizing user behavior regardless of health outcomes. The study by Alonso-Fernández et al [23] used the metrics collected during gameplay to predict posttest outcomes using machine-learning algorithms. The study by Loh et al [24] examined the course of actions of players and used several similarity measures to compare players, aiming to differentiate them efficiently and create gaming profiles (distinguishing among fulfillers, explorers, and quitters) in SG. This indicates the potential of using in-game metrics to model different outcomes, depending on what needs to be evaluated.

On the health assessment front, the evaluation and validation of in-game metrics as assessment tools are performed either in comparison with a clinical diagnosis or a validated assessment test [25]. Cognitive measures in game-like interfaces can contribute to the early detection of neurological disease [26]. The study by Bang et al [27] used an index calculated from in-game metrics to detect children with heterogeneous developmental disabilities, whereas the study by Kim et al [28] evaluated the use of kinetic variables from the interactions with an SG as a digital biomarker for developmental disabilities.

SG targeting the improvement of the physical capacity of the player, as the one addressed in this study, are called exergames and can open up a new category of outcome assessment. Exergames are a promising tool for measuring and assessing unobtrusively physical health [19,21] and focus mainly on fall risk assessment by correlating typical in-game metrics of exergames, such as movement time and response time, with a test battery or standardized assessment tests of fall risk [29]. The study by Aguilar et al [30] assessed the effect of 6 weeks of unsupervised home-based SG intervention on dynamic postural control. They used generalized linear models and classification algorithms to estimate the probability that the body movements recorded by Kinect belonged to a participant older than 60 years, and the objective was to distinguish between younger and older participants. However, no further insights

XSL•FO
RenderX

were provided regarding the well-being state of the participants. The study by Pirovano et al [25] developed an SG solution to support rehabilitation at home. Their solution combined fuzzy-based monitoring and in-game adaptation to capture the knowledge of the clinician and provide real-time feedback during exercise. This feedback was used for adaptation of the gameplay but not for assessment, although it might have the potential to capture insights for the user movement and rehabilitation process.

Moreover, a review of existing literature has shown that there is a strong interest from the research community in the use of SG, and especially in-game metrics, as psychometric tools and indicators [31]. The study by Valladares-Rodríguez et al [31] identified research issues related to the development of SG for use in neuropsychological evaluation, proving its potential as an alternative to conventional neuropsychological examinations. However, it is pointed out that more research is needed on their reliability and validity for their application in daily clinical practice [32]. In addition, it is necessary to address the risk of investing in technical features that could potentially affect the reliability of the game. For example, to make it so attractive that, during the interaction, the game provokes the respective feature that is called to measure, thus intertwining the purpose of enhancing a feature with that of its measurement [33]. At the same time, they provide many opportunities to enhance the reliability of evaluation processes. In-game metrics can provide information associated not only with the performance outcome of a specific test but also with the processes during the test.

Study Objectives

This study investigated the possibility of using in-game metrics from exergame interactions as digital biomarkers for well-being. The digital biomarkers investigated are produced from in-game metrics analysis and aim to support the creation of health and well-being profile groups, as well as assess the physical and cognitive state of a user from gameplay without ex situ data. The in-game metrics collected during the interaction of patients with Parkinson disease with the webFitForAll platform [34,35] were used as a case study to validate the study objectives.

Methods

Overview

The main analysis included the clustering of participants using the neuropsychological and physical clinical assessment tests that define the ground truth. The clustering was evaluated to select the best grouping of participants. A classification method was used to predict the group to which each participant belonged, using in-game metrics as features. The correlation of each in-game metric with each clinical assessment test was examined to identify those that had a higher separation value. The methodology followed does not rely on the metric choice or any prior knowledge of in-game metrics. We are trying to be metric agnostic, meaning that we are trying to find an explanation for how the data correlations with the clinical assessment tests are ruled out.

webFitForAll Platform

This study analyzed the data captured during the interaction with the webFitForAll platform [20]. webFitForAll is a web-based platform that provides SG for exercising that are specially designed and tested for older adults and vulnerable populations. The users interact with the game using body movements captured with 3D depth sensor controllers. There are additional games on the platform that are controlled by other types of user interfaces, such as touch screens and voice control. This study focuses on the exergames of the platform, which are controlled by body movements to obtain unified results. The games used in this study were as follows: (1) fishing, (2) kinematic orchestra, (3) picking citrus fruits, and (4) retraining in eating behavior [36]. The games were designed to address specific gait, balance, and exercising needs of patients with for Parkinson disease using a participatory design methodology [36].

In the fishing game, the user's body posture is translated into the direction and velocity of a digital boat. Leaning forward will result in a forward movement of the boat with acceleration proportional to the body inclination. The goal is to collect as many fish as they can, in a specific period, by driving the boat toward the fish, avoiding at the same time the obstacles (rocks and sharks) and counterbalance the wind that might alter their direction [37].

In the kinematic orchestra game, the user tries to associate a specific group of notes with hand gestures such as moving the left or right hand up, down, or cyclical, or lower body movements such as raising the right or left leg. The note groups are presented sequentially, and the user is given a specific period to recognize the group, match it with a specific movement, and perform the movement [37].

In the picking citrus fruits game, the user navigates in a virtual environment by walking on spot. Specific instructions on picking and putting down fruits are presented on the screen regarding the sequence of actions the user needs to perform. The user is picking fruits by moving either the left or right hand. The climbing is simulated by walking on spot. The goal is to pick as many fruits as possible in a specific period [37].

In the retraining of the eating behavior game, the user faces the screen, either seated or standing, and pretends to hold a spoon or fork. An avatar is presented on the screen and shows the correct frequency of movement. The user moves their hand to imitate the movement of bringing the spoon in the mouth. The game monitors the movement and correlates it with the correct movement presented on the screen. Every time the user maintains the correct frequency, they earn a point. The game lasted for a specific period [37].

Data Set

In-Game Metrics

During participant interaction with the games, the system manually captures metrics that are representative and can provide insights for each game. For every game, different in-game metrics correspond to different measures. Each in-game metric is described in Textbox 1.

XSL•FO
RenderX

Textbox 1. In-game metrics.

Fishing

- Score: this in-game metric captures how many fish the user collects during a specific time of gameplay. If a user runs into obstacles (sharks or rocks), the score is reduced. Therefore, the score is a combination of moving toward the goal (fish) and moving away from obstacles (sharks and rocks).
- Goal time: it is the duration between the time point when the target fish is presented in the screen until the time point that the user captures the fish. This duration was calculated for each fish caught by the user. For each session (Si), the in-game metric is a sequence of values representing the time for each reached goal.

Kinematic orchestra

- Score: this in-game metric measures the number of movements correctly performed within a limited period. The game recognizes whether the user has performed the correct movement based on the matching between note groups and movements, and that the user was able to react quickly enough.
- Goal time: this is the duration between the time point when the target note group is presented on the screen until the time point when the user performs the right movement. This duration was calculated only for movements performed correctly. For each session (Si), the in-game metric is a sequence of values representing the time for each reached goal.

Picking citrus fruits

- Score: this in-game metric captures the number of fruits the user collects during a specific time of gameplay.
- Goal time: this is the duration between the time point when the targeted fruit becomes highlighted and the time point that the user "catches" the fruit. This duration is calculated for each fruit that the user catches. For each session (Si), the in-game metric is a sequence of values representing the time for each reached goal.

Retraining of eating behavior

- Score: this in-game metric captures the number of correct movements that the user performs during a specific time of gameplay. Correct movement is considered to occur almost simultaneously with the avatar presented on the screen. The game allows a specific time window that is sufficiently small to consider that the movement is performed with the same frequency.
- Goal time: this is the difference between the time point at which the avatar performs the movement and the time at which the user performs the movement. For each session (Si), the in-game metric is a sequence of values representing the time for each reached goal.

All data collected during the sessions were stored in an SQL database, pseudoanonymized, and password-protected for each participant. The data set was retrieved from the database for offline analysis and fully anonymized for this study.

Clinical Neuropsychological and Physical Assessment Tests

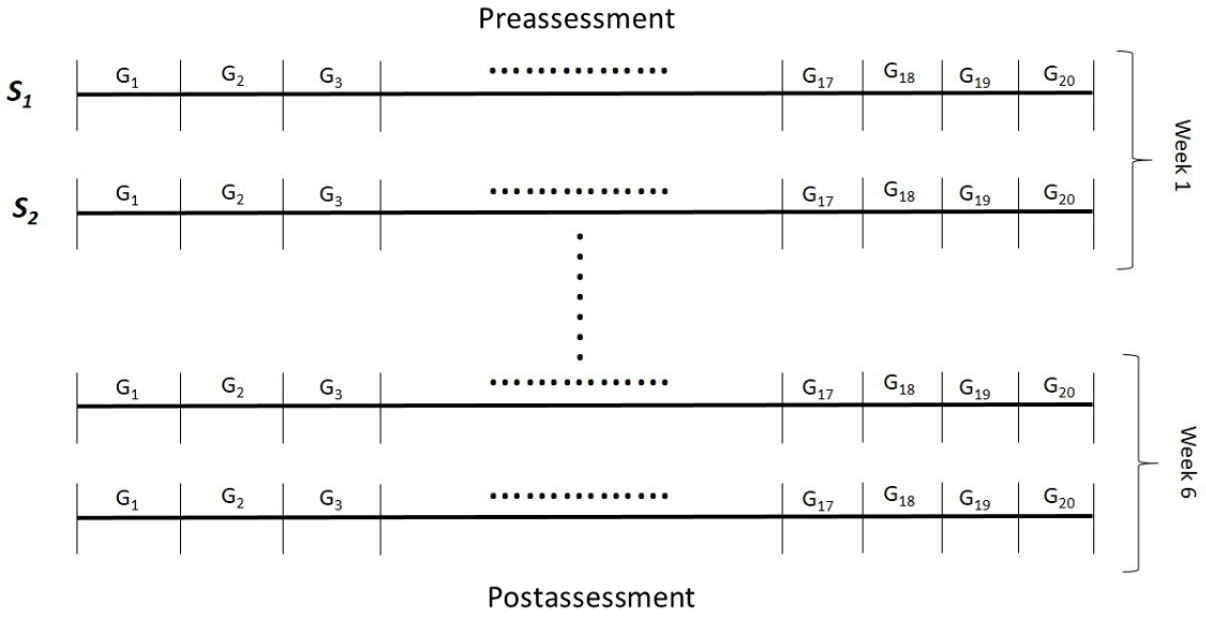
Before and after every sequence of intervention, in a time window of no longer than 1-week, clinical neuropsychological and physical assessment tests were administered to each participant. These tests were performed by professionals, psychologists, and physical educators. They were considered the ground truth for each participant's cognitive and physical state before and after the intervention. The administered tests were carefully selected to depict all the domains that were influenced by the SG, as well as the domains that were mostly influenced by Parkinson disease. The selected tests assess various levels of physical status as well as cognitive impairment and meet specific criteria such as validity, reliability, and objectivity. The administered tests were the Single-Leg-Stance Test [38], Berg Balance Scale [39], Short Physical Performance Battery [40], Community Balance and Mobility Scale (CB&M), Senior Fitness Test (Fullerton Fitness Test) [41], BMI [42], Performance-oriented Mobility Assessment [43], 10 Meter walk, Instrumental Activities of Daily Living Scale [44], 8-item Parkinson's Disease Questionnaire [45], Fall Risk Assessment [46], Dementia Rating Scale [47], and Symbol Digit Modalities Test (Symbol) [48].

Protocol

Every participant enrolled in the study had to attend at least 16 sessions not to be considered a dropout. Every session (Si) comprised the same sequence of games (Gi), including 20 games in total. The sessions were performed twice a week on predefined days. A participant could be reenrolled in a study for a follow-up series of interventions. However, the data from follow-up sessions were not considered a unified participant study if they were performed in a period of more than 1 week from the last session. Figure 1 presents a visualization of the protocol performed by each participant. A participant study was defined as one in which data were collected from sessions performed continuously, composed the same game sequence in every session, and came from a single participant.

XSL•FO
RenderX

Figure 1. Visualization of a participant study. G1 represents the sequence of games.



From the aforementioned sequence of games (Gi), 4 were used for the analysis (fishing, kinematic orchestra, picking citrus fruits, and retraining in eating behavior). These correspond to exergames and are captured in a unified manner using depth sensor cameras. These are the ones that require body movement interactions with the game and were suggested by health care professionals as the most indicative for capturing the patient's condition and the most commonly appearing Parkinson disease symptoms. The protocol also included preclinical and postclinical neuropsychological and physical assessments using standardized questionnaires, as presented in the previous section. The participants were also engaged in 1 testing session before the actual intervention period to familiarize themselves with the games and eliminate the effect of nonrepresentative game measures owing to misunderstandings in the first intervention.

Participants

The experimental data set that was used for testing the approach in this study consists of gaming sessions that took place within the i-Prognosis H2020 project [49] within day care centers of the "Northern Greece Association of Parkinson's Disease Patients and Friends." A total of 13 participants, all diagnosed with Parkinson disease, with a mean age of 64.5 (SD 9.3) years, participated in the study protocol. All participants signed an informed consent form, and no financial incentives were provided to them. The participants interacted with the webFitForAll platform twice a week in sessions of 60 minutes each. Each participant performed a mean number of 25.9 (SD 5.4) sessions. Clinical assessment tests were administered within a week before entering the study (before the assessment) and within 1 week after completing the series of intervention sessions (after the assessment).

Analysis Methodology

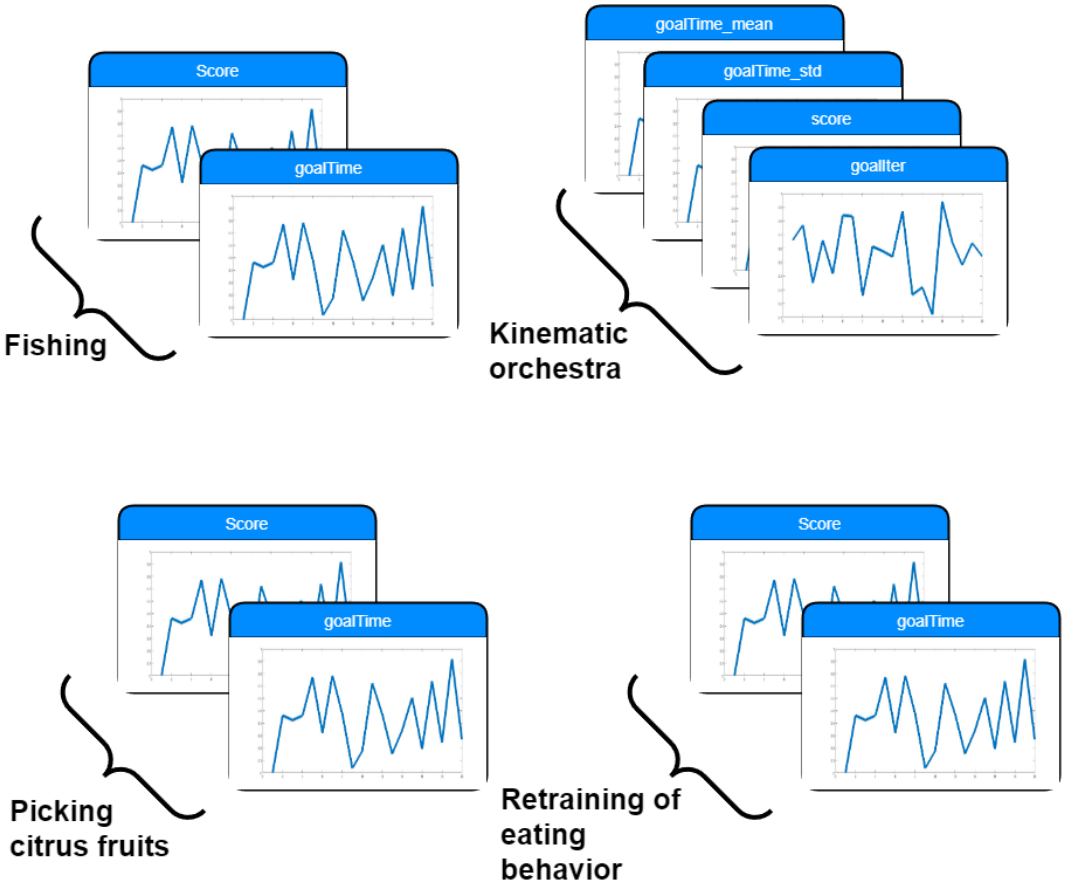
Preprocess

The mean and SD values were calculated for each in-game metric per session. The result for each user is a time series for each in-game metric, with every point in the time series representing the value for 1 session. Each user played various games and different in-game metrics were captured for each game. A visual representation of the data set collected for each user is shown in Figure 2.

After collecting all the data, the outliers for each in-game metric were found and removed using the IQR based on the quartile method for the detection of outliers [50]. IQR is the range between the median of the upper and lower halves of the data. A total of 4 quartiles were computed for each in-game metric series, and the IQR was calculated as IQR=Q3-Q1. A point in an in-game metric was considered an outlier if its value exceeded the value of 1.5×IQR and was then removed from the data set. In addition, we removed extreme values that were produced when the human skeleton was not detected properly because of the hardware resolution and monitoring conditions. Some extreme values were also produced by software bugs that were subsequently identified using the log files of the system.

XSL•FO
RenderX

Figure 2. Visual representation of mock data set representing a participant study.



The next step was to extract the features from each in-game metric time series. The mean and SD values of the time series were selected as features. The first 5 values of each in-game metric time series, henceforth referred to as PRE in-game metric data, were calculated for each in-game metric feature. The first 5 sessions were considered to better reflect the physical and mental state of the participants before the beginning of the intervention, absorbing any artifacts from the games learning procedure. Considering only 1 game point in the analysis can affect the reliability of results.

Clustering

A set of neuropsychological and physical assessment tests was used as the ground truth for separating participants into groups with better or worse physical and cognitive states using clustering analysis. The hierarchical agglomerative clustering (HAC) algorithm was used for clustering. In HAC, each observation is initially considered a separate class. The algorithm chooses clusters that are most similar to each other and merges them into 1 cluster, which continues until all objects are merged into a single cluster.

Classification

A classification method was used to predict the group each participant belongs to using, as features, the in-game metrics. Each feature set consists of all in-game metric values following the feature extraction procedure described earlier. The decision tree classifier was selected as the classification method because the internal decision-making logic is clear and transparent, which is not the case for black box-type algorithms such as neural networks. Decision tree classifiers are fast to train and allow the capture of descriptive decision-making knowledge that can help us interpret results. The decision rules produced by the decision tree classifier can also be exploited in the design and interpretation of other in-game metrics. The selection measure used in this study is the Gini index, which measures the probability that a specific variable is incorrectly classified when its class is randomly chosen.

The leave-one-out method was used to mitigate the low volume of data. In this method, the data set was first separated into m number of data sets each containing 1 feature vector for 1 participant. In each iteration, 1 feature vector was kept for testing, whereas the other m-1 feature vectors were used for training the model. When all feature vectors have been used 1 time for testing, the process is complete. Thus, for every feature vector, there is an assigned class, which is the predicted class.

Analyzing Each In-Game Metric Separately

The aim of this step is to identify the contribution of each in-game metric to the prediction of the well-being status that corresponds to each assessment test. To do so, a feature vector containing the mean value of all sessions for each in-game metric was calculated for each user. In addition, the clinical assessment test feature vector was calculated, which corresponded to the different states of well-being. This vector consisted of the mean value for each preassessment and postassessment test. The Pearson correlation coefficient between the in-game metrics and clinical assessment tests was calculated. The results were considered to specify which in-game metrics

XSL•FO
RenderX

could predict which assessment test. Metrics with a strong correlation (high Pearson correlation coefficients) were selected for further analysis.

Representing each time series with a single value lacks information about the progression of values and how they change over time, which might be present in the whole time series. To avoid this and consider the evolution of each participant throughout the sessions, the dynamic time warping (DTW) [51] method was used. DTW is a method used in time series analysis to measure the similarity of 2 time series, even if they have differences in speed. For example, 2 time series can have the same form, but vary in time length and data points. The values that demonstrated a strong correlation in the previously described step were used to calculate the distance matrix using the DTW method. This distance matrix was then used as the input for HAC analysis.

Ethics Approval

This study was approved by the School of Medicine Bioethics Committee (protocol number 4.123, 17/7/2019).

Results

First, we present the results of the general analysis considering all games and their in-game metrics, as well as all clinical assessment tests. Then, the results presented focus on the picking citrus fruits score in-game metric, which has the highest predictive value according to the general results.

Clustering

The HAC clustering algorithm using the Euclidean distance metric and ward linkage resulted in the following groups:

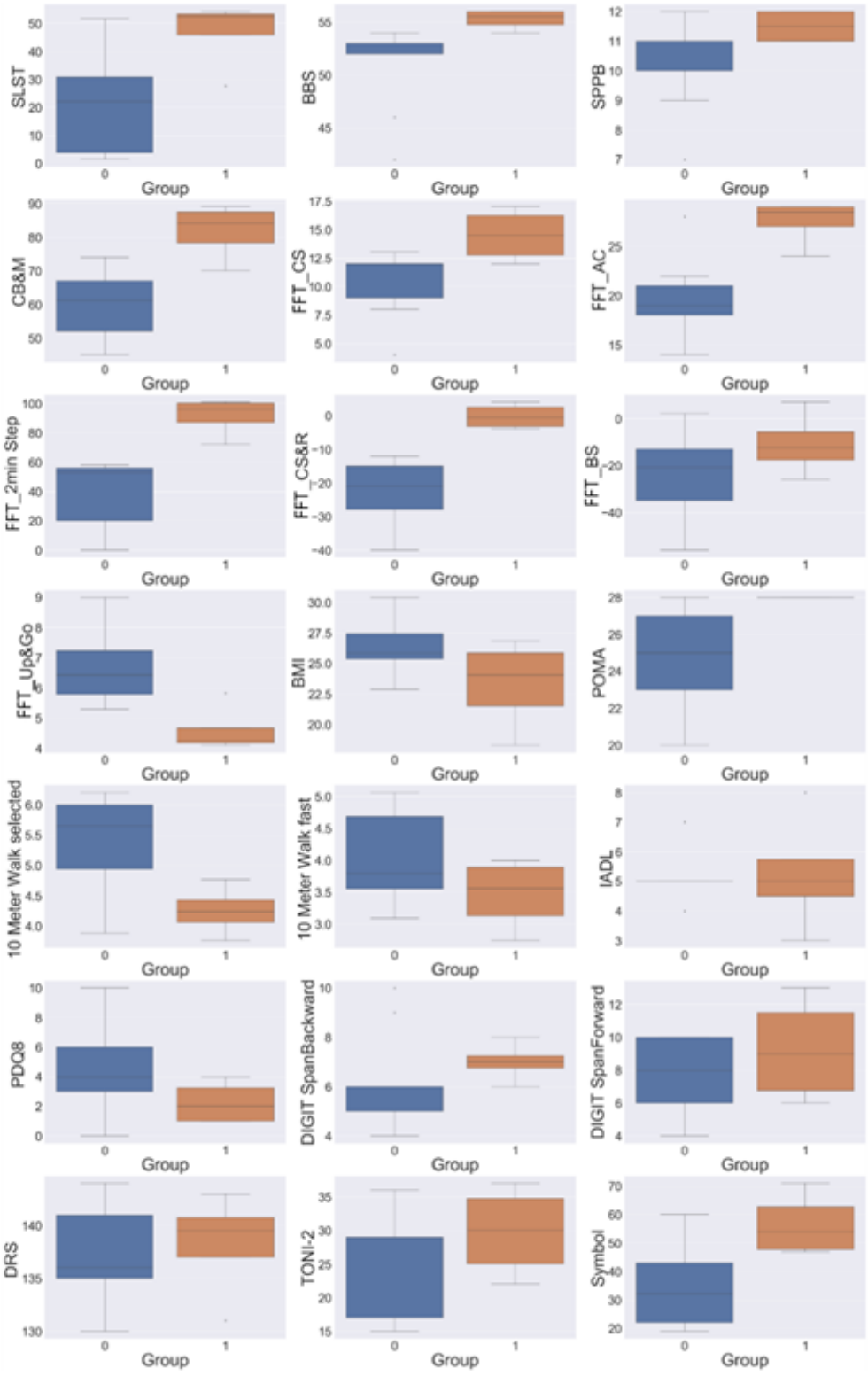
- Group 0: p#2, p#6, p#8, p#9
- Group 1: p#0, p#1, p#3, p#4, p#5, p#7, p#10, p#11, p#12

Figure 3 presents boxplots of the clinical assessment tests for the 2 different groups of participants that were formed after clustering. In most cases, the 2 formed groups demonstrated a large difference in the distribution of values, which is an indication of good intercluster differences.

In general, group 0 scored lower range of values compared with group 1 in most tests. The image is different in Fullerton Fitness Test (FFT) foot up and go, BMI, and 10 Meter Walk, but these are tests that show higher values indicate lower capacity. Thus, we can safely conclude that group 0 included participants with lower physical and cognitive capacities, whereas group 1 included participants with higher physical and cognitive capacities.

XSL•FO
RenderX

Figure 3. Clinical assessment test values for the 2 different groups formed after clustering. BBS: Berg Balance Scale; CB&M: Community Balance and Mobility; FFT: Fullerton Fitness Test; FFT_AC: Fullerton Fitness Test arm curl; FFT_BS: Fullerton Fitness Test back scratch; IADL: Instrumental Activities of Daily Living; PDQ-8: 8-item Parkinson's Disease Questionnaire; POMA: Performance-Oriented Mobility Assessment; SPPB: Short Physical Performance Battery; TONI-2: Test of Nonverbal Intelligence, 2nd edition.



Classification

The decision tree classifier (criterion=Gini) was evaluated using the leave-one-out method. The goal was to predict if a user belongs to group 0, indicating lower physical and cognitive capacity or in group 1, indicating higher physical and cognitive capacity using information coming only from in-game metrics. The mean values of the initial measurements (PRE) of the in-game metrics were used as features for the prediction. A high classification accuracy (0.846) shows the capacity of in-game metrics to distinguish users with comparable discriminant values to the clinical assessment tests. The confusion matrix and evaluation metrics are presented in Table 1.

XSL•FO
RenderX

Table 1. Confusion matrix and results from classification.

True label	Predicted label[a]	
	Zero	One
One	• True negative • 8[b] • 61.54%	• False Positive • 1 • 7.69%
Zero	• False negative • 1 • 7.69%	• True Positive • 3 • 23.08%

[a]Accuracy 0.846; recall 0.75; precision 0.75; F_1-score=0.75.

[b]The absolute numbers are the instances that were true negative, false positive, etc, and the percentage of instances to the total number of instances.

The model also demonstrated high recall, precision, and F_1-score values, which further supports the sensitivity of the model for detecting participants with both higher and lower levels of physical and cognitive capacity.

Participant p#12 was falsely assigned to group 1 by the classifier, whereas participant p#2 was falsely assigned to group 0.

Analyzing Each In-Game Metric Separately

The Pearson correlation coefficient of the in-game metrics series with the clinical assessment tests is presented in Figure 4. Only the correlations with absolute values higher than 0.6 are presented. In-game metrics with a high correlation with specific test values can be used to build a model that predicts the exact value of the assessment test and not just a general class that separates the group of participants into higher and lower physical and cognitive capacity.

From Figure 4, we conclude that the picking citrus fruits score in-game metric has a higher degree of correlation with most clinical assessment tests. Thus, this in-game metric is the most reliable for further investigation of its correlation with each clinical assessment test separately. Although the picking citrus fruits score has a strong correlation with various assessment tests, this in-game metric alone cannot be used to separate participants into 2 categories based on their physical and cognitive status.

The time series for each participant for the picking citrus fruits score in-game metric was used to calculate the distance matrix using the DTW method. This distance matrix, using HAC with complete linkage, resulted in 3 clusters based on their performance in the specific metric (picking citrus fruits score):

- Group 1: p#9, p#6, p#8
- Group 2: p#3, p#5
- Group 3: p#7, p#12, p#11, p#0, p#4, p#2, p#1, p#10

To further assess the results of clustering, we present the CB&M, the FFT arm curl, and the FFT 8-foot up and go mean prevalues and postvalues in correlation with the mean picking citrus fruits score in-game metric, which are the tests with higher correlation values, as shown in Figure 5. The different identified groups are presented with different colors.

All 3 clinical assessment tests demonstrated a strong correlation with the in-game metric (r=0.82, r=−0.87, and r=0.85) and clear separation of the 3 clusters created using the whole time series. Group 0 corresponds to participants with lower capacity and performance, group 1 corresponds to medium performance, and group 2 includes the best players with higher physical capacity. Participants p#6, p#8, and p#9 were assigned to the clusters that correspond to the “good” participants in both the aggregated study and considering only 1 in-game metric. Participant p#0 was close to the lower capacity group in all cases but was clustered in the medium performance group based on the picking citrus fruits game score that the participant achieved over time. Participant p#0 was assigned to the higher-capacity group in the clustering, which shows the importance of considering a combination of in-game metrics for the assessment. It is important to point out that the 3 tests with higher correlation values assess physical capacity, which indicates that it is important to include other in-game metrics to assess cognitive capacity as well.

XSL•FO
RenderX

Figure 4. Pearson correlation of in-game metrics series with assessment tests. BBS: Berg Balance Scale; CB&M: Community Balance and Mobility; FFT: Fullerton Fitness Test; IADL: Instrumental Activities of Daily Living; POMA: Performance-Oriented Mobility Assessment; SPPB: Short Physical Performance Battery.

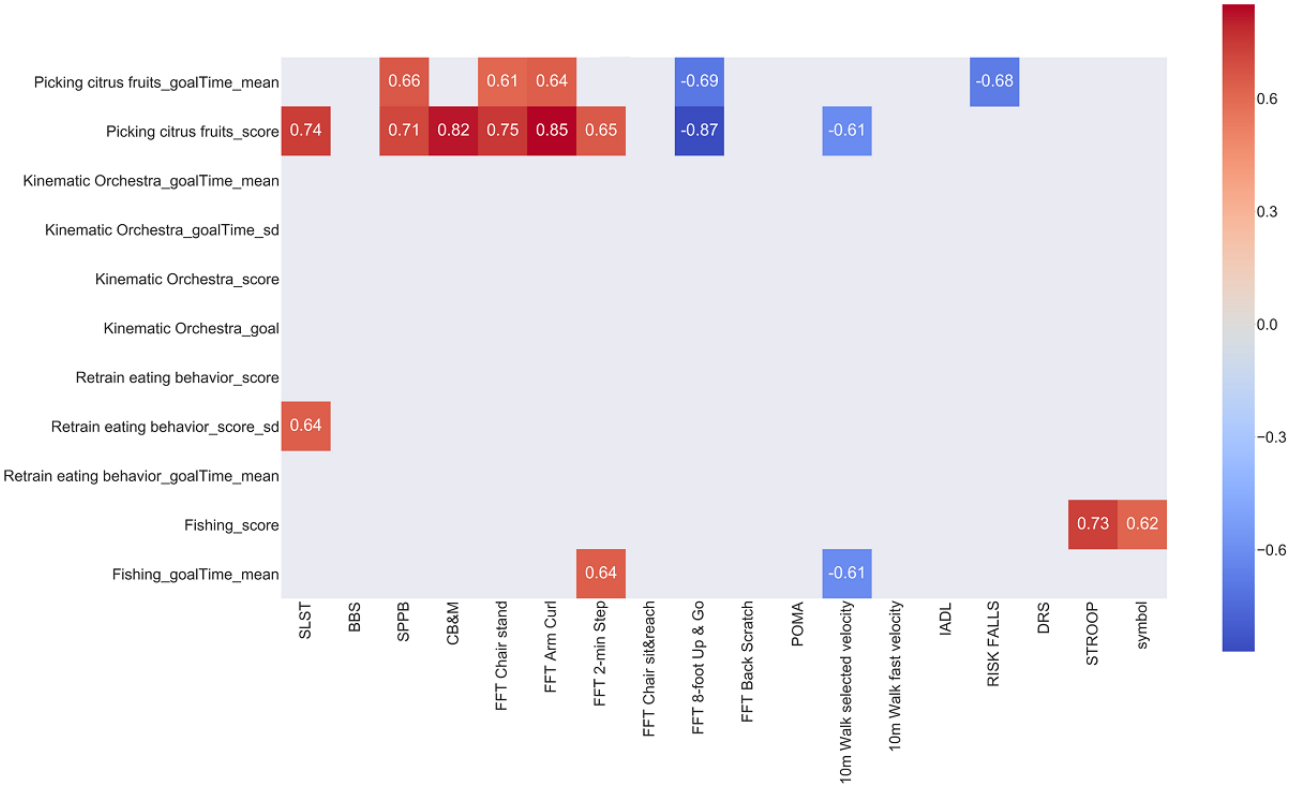
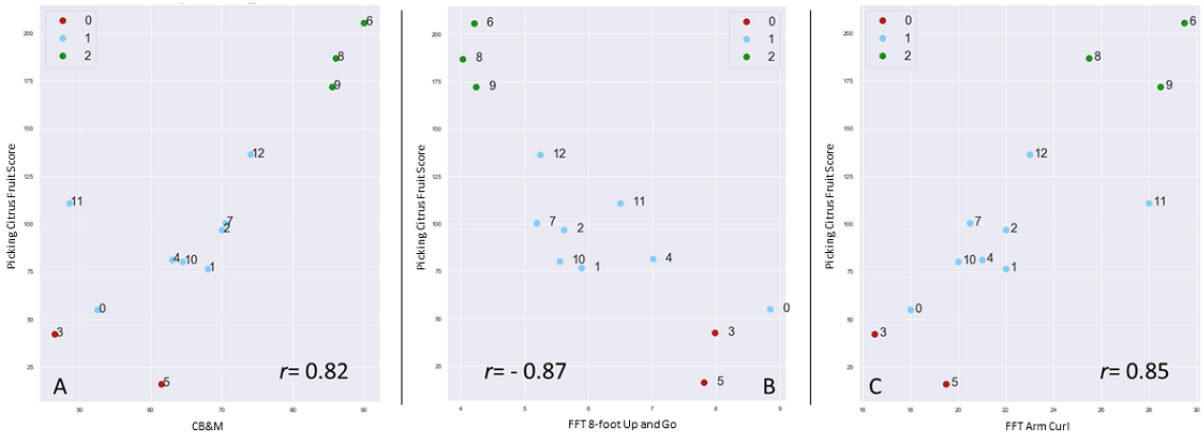


Figure 5. Community Balance and Mobility test versus picking citrus fruits game score (right), Fullerton Fitness Test (FFT) 8-foot up and go test versus picking citrus fruits game score (center), and 10 FFT arm curl test versus picking citrus fruits game score (left). The identified groups are presented with different colors. CB&M: Community Balance and Mobility; FFT: Fullerton Fitness Test.



Discussion

Principal Findings

This study explored the possibility of using in-game metrics as digital biomarkers to gain meaningful insights into participants' physical and cognitive states. The analysis was performed in two different steps: (1) considering the whole data set of clinical assessment tests to compute a baseline and then classifying participants based on in-game metrics and (2) evaluating the value of a single in-game metric to predict the participants' physical state.

The analysis returned very promising results, achieving an 85% accuracy in distinguishing the 2 groups of participants based on their performance in the preneuropsychological and physical assessment tests (group 0 indicates lower physical and cognitive capacity, whereas group 1 indicates higher physical and cognitive capacity). Considering the classification of the participants from in-game metrics, a professional can obtain an idea for the participant profile based on the characteristics of each group in the corresponding cognitive and physical assessment (Figure 4). The proposed methodology can be generalized for further use in other types of SG and in-game

XSL•FO
RenderX

metrics using different standard tests as the ground truth for each user, as it is metric agnostic.

Although the use of a combination of in-game metrics as a digital biomarker for characterizing well-being returned promising results, the separate use of each in-game metric can provide more detailed information for specific physical and cognitive capacities. Picking citrus fruits score in-game metric achieved a high correlation with multiple physical assessments and can be further considered as a digital biomarker for characterizing physical status. The picking citrus fruits game that demonstrated the highest descriptive value, targets improvement of motor skills, synchronization, and balance skills [52]. It shows high correlations with physical tests, such as FFT and CB&M, which provide insights into the physical state of the person and the movement of the lower body. Achieving a higher score in the picking citrus fruits game can be explained by the fact that a player with higher stability and strength in the lower body has higher confidence and performs the exercise with precision and speed. This translates to higher clinical assessment scores.

Insights From Professionals

A total of 4 professional experts in the health and well-being domain assisted the participants in their day-to-day interactions with the games and performed clinical assessments. The professionals evaluated the participants' performance, overall capacity, and value of the proposed games. Professionals considered the fishing and picking citrus fruits games as games that can more effectively separate participants into 2 groups. This supports the results of this study, as the 2 games yielded higher correlations with the ground truth assessment test.

Participants p#6, p#8, and p#9 were considered as the players with the higher capacity and performance by professionals, which is also reflected in the results of the current analysis. These players were always classified in the group of "better" cognitive and physical state. The professionals working with p#12 commented that the participant was not very concentrated during the sessions and used to talk frequently. This may explain the false classification results and incoherent outcomes between the 2 analyses. Participant p#0 was assigned to group 1 based on the picking citrus fruits in-game metric, although it had low scores in physical assessment tests. Professionals working with the participant commented that besides the low physical capacity, the participant had tried a lot and showed great progress. This explains why the DTW distance score was lower for participants in group 1, indicating a tendency to improve, and hence, assign the participant to that group.

In the case of a system with both interventional and assessment capabilities, such as webFitForAll, the value for the participants could be 2-fold. First, delay in cognitive decline onset as physical exercise is a preventive intervention for cognitive decline [53]. Second, early detection of cognitive and physical decline symptoms would provide the opportunity for the early administration of available treatments when interventions are more effective [34,54].

The presented methodology could be useful in categorizing players with very high and poor performance. The transition from one group to another throughout the course of the intervention sessions can be an alarm for further evaluation. The first set of sessions can also be exploited to collect data to guide the design of individualized interventions and specify areas of difficulty and behavioral response patterns to the skills being tested.

Comparison With Previous Studies

There is an increasing body of research that investigates the complementarity of digital tools for measuring health, well-being, and clinical outcomes, along with existing methods. Similar studies have investigated the use of 3D depth sensor cameras (Microsoft Kinect) in game design to identify patients with spinal muscular atrophy and healthy controls [55]. Similar to our results, they identified some digital biomarkers that can detect differences (eg, hand velocity), whereas other minor differences in functioning cannot be detected. The study by Gielis et al [56] used a combination of data produced by a casual card game as digital biomarkers to distinguish mild cognitive impairment from healthy participants. Their results were similar to ours (accuracy 0.792), but a direct comparison of the 2 studies could not be performed. However, both studies focused on the suitability of analyzing in-game metrics for characterizing participants' health and well-being.

In-game metrics cannot substitute for the use of clinical neuropsychological and physical evaluation, but can be used as an auxiliary method of amplifying and cross-referencing results from the traditional method of evaluation. Beyond SG, recent studies present computerized forms of cognitive assessment tests where the results are compared against the corresponding pen and paper tests, which exhibit a strong correlation [26]. These game-like screening tests focus mainly on the assessment of the player and not on the intervention, thereby confusing the term "game" with the term "computerized test." In addition, a number of computerized cognitive screening tests have been evaluated with respect to their sensitivity and specificity with high accuracy [57-59]. Each SG consists of a multifactorial stimulus and improvement features, whereas assessment tests provide a structured evaluation method that specifically targets 1 factor. In-game metrics can provide detailed and rich data on processes and the progress of the participant, which can be used for evaluation over time and individually for each participant.

Although digital biomarkers from SG show significant discriminative value, a more in-depth contextual analysis is required per case. A better understanding of each target group's capabilities and particularities can lead to further adaptation and selection of measures.

Limitations

Some of the limitations of the study are the small number of participants (N=13) and individual factors, such as partial symptomatology and the course of Parkinson disease, that significantly differentiated the conditions for each of the participants and because of the small sample size as variables bear great weight in the influence of the results. Owing to the small sample size, the above method of dividing into groups of users with good performance and those with poor performance

XSL•FO
RenderX

is not very sensitive in locating players with more specific performance profiles and unstable performance.

In addition, most games were played with the help of a facilitator. The facilitator played a supporting role, but sometimes their comments could bias the in-game metric results. This research methodology could be applied to larger samples and provide safer results, as the influence of individual endogenous and environmental factors would be reduced. Factors such as hesitation to participate, stress, or preexisting depression can be assessed at the beginning of the intervention and can be used to weigh the individual results as they can affect the engagement and effort put by each participant.

Finally, the SG were initially designed as interventions and did not aim to assess the participants. However, as their value as screening methods becomes apparent, a redesign in that direction could support their valorization as decision-support tools.

Conclusions

SG have recently been used as a reach source of unobtrusively captured information about the user that can drive the creation of digital biomarkers for assessing health and well-being. This study explores the use of the webFitForAll platform, which collects in-game metrics from user movement during gameplay, to identify different user profiles compared with a baseline created by clinical assessment tests. The results are promising and can boost the analysis for improving in-game metrics to obtain more detailed results. More in-game metrics can be gathered during the analysis, specifically targeting the prediction of assessment tests.

Acknowledgments

The authors would like to thank the facilitators Ioanna Dratsiou, Maria Metaxa, Foteini Dratsiou, Maria Karagianni, and Sotiria Gylou, who supported the participants throughout the serious games session and shared meaningful, individualized insights for performance and gameplay. The authors would also like to thank the participants during the intervention sessions.

This work was supported in part by the CAPTAIN Horizon 2020 project (grant 769830), the i-Prognosis Horizon 2020 project (grant 690494), and the SHAPES Horizon 2020 project (grant 857159).

Conflicts of Interest

None declared.

References

1. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertain Comput 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
2. Constitution of WHO: principles. World Health Organization. 2018. URL: https://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf [accessed 2021-09-01]
3. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry 2017 Jan 18;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
4. Nguyen TT, Ishmatova D, Tapanainen T, Liukkonen TN, Katajapuu N, Makila T, et al. Impact of serious games on health and well-being of elderly: a systematic review. In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: HICSS '17; January 4-7, 2017; Waikoloa, HI, USA. [doi: 10.24251/hicss.2017.447]
5. Calderón A, Ruiz M. A systematic literature review on serious games evaluation: an application to software project management. Comput Educ 2015 Sep;87:396-422. [doi: 10.1016/j.compedu.2015.07.011]
6. Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of serious games: an overview. Adv Human Comput Interact 2013 Jan;2013:1-11. [doi: 10.1155/2013/136864]
7. Loh CS, Sheng Y. Measuring expert performance for serious games analytics: from data to insights. In: Loh CS, Sheng Y, Ifenthaler D, editors. Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement. Cham, Switzerland: Springer; 2015:101-134.
8. Ram N, Brinberg M, Pincus AL, Conroy DE. The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. Res Hum Dev 2017;14(3):253-270 [FREE Full text] [doi: 10.1080/15427609.2017.1340052] [Medline: 30613195]
9. Friehs MA, Dechant M, Vedress S, Frings C, Mandryk RL. Effective gamification of the stop-signal task: two controlled laboratory experiments. JMIR Serious Games 2020 Oct 08;8(3):e17810 [FREE Full text] [doi: 10.2196/17810] [Medline: 32897233]
10. Sajjadi P, Ewais A, De Troyer O. Individualization in serious games: a systematic review of the literature on the aspects of the players to adapt to. Entertain Comput 2022 Mar;41:100468. [doi: 10.1016/j.entcom.2021.100468]
11. Serrano-Laguna Á, Martínez-Ortiz I, Haag J, Regan D, Johnson A, Fernández-Manjón B. Applying standards to systematize learning analytics in serious games. Comput Stand Interfaces 2017 Feb;50:116-123. [doi: 10.1016/j.csi.2016.09.014]

XSL•FO
RenderX

12. Alves T, Gama S, Melo FS. Flow adaptation in serious games for health. In: Proceedings of the IEEE 6th International Conference on Serious Games and Applications for Health. 2018 Presented at: SeGAH '18; May 16-18, 2018; Vienna, Austria p. 1-8. [doi: 10.1109/segah.2018.8401382]
13. Borghese NA, Pirovano M, Lanzi PL, Wüest S, de Bruin ED. Computational intelligence and game design for effective at-home stroke rehabilitation. Games Health J 2013 May;2(2):81-88 [FREE Full text] [doi: 10.1089/g4h.2012.0073] [Medline: 24761321]
14. Froschauer J, Seidel I, Gärtner M, Berger H, Merkl D. Design and evaluation of a serious game for immersive cultural training. In: Proceedings of the 16th International Conference on Virtual Systems and Multimedia. 2010 Presented at: VSMM '10; October 20-23, 2010; Seoul, South Korea p. 253-260. [doi: 10.1109/vsmm.2010.5665978]
15. Swarz J, Ousley A, Magro A, Rienzo M, Burns D, Lindsey A, et al. CancerSpace: a simulation-based game for improving cancer-screening rates. IEEE Comput Graph Appl 2010;30(1):90-94. [doi: 10.1109/MCG.2010.4] [Medline: 24807098]
16. Bellotti F, Berta R, De Gloria A. Designing effective serious games: opportunities and challenges for research. Int J Emerg Technol Learn 2010 Nov 22;5(SI3):22-35. [doi: 10.3991/ijet.v5s3.1500]
17. Bamparopoulos G, Konstantinidis E, Bratsas C, Bamidis PD. Towards exergaming commons: composing the exergame ontology for publishing open game data. J Biomed Semantics 2016 Feb 9;7:4 [FREE Full text] [doi: 10.1186/s13326-016-0046-4] [Medline: 26865947]
18. Mandryk RL, Birk MV. The potential of game-based digital biomarkers for modeling mental health. JMIR Ment Health 2019 May 23;6(4):e13485 [FREE Full text] [doi: 10.2196/13485] [Medline: 31012857]
19. Staiano AE, Calvert SL. The promise of exergames as tools to measure physical health. Entertain Comput 2011 Jan 01;2(1):17-21 [FREE Full text] [doi: 10.1016/j.entcom.2011.03.008] [Medline: 23378860]
20. Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD. Design, implementation, and wide pilot deployment of FitForAll: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE J Biomed Health Inform 2016 Jan;20(1):189-200. [doi: 10.1109/JBHI.2014.2378814] [Medline: 26731797]
21. Konstantinidis EI, Bamidis PD, Billis A, Kartsidis P, Petsani D, Papageorgiou SG. Physical training in-game metrics for cognitive assessment: evidence from extended trials with the Fitforall exergaming platform. Sensors (Basel) 2021 Aug 26;21(17):5756 [FREE Full text] [doi: 10.3390/s21175756] [Medline: 34502647]
22. Anagnostopoulou A, Styliadis C, Kartsidis P, Romanopoulou E, Zilidou V, Karali C, et al. Computerized physical and cognitive training improves the functional architecture of the brain in adults with Down syndrome: a network science EEG study. Netw Neurosci 2021 Mar 1;5(1):274-294 [FREE Full text] [doi: 10.1162/netn_a_00177] [Medline: 33688615]
23. Alonso-Fernández C, Martínez-Ortiz I, Caballero R, Freire M, Fernández-Manjón B. Predicting students' knowledge after playing a serious game based on learning analytics data: a case study. J Comput Assist Learn 2019 Dec 11;36(3):350-358. [doi: 10.1111/jcal.12405]
24. Loh CS, Li I, Sheng Y. Comparison of similarity measures to differentiate players' actions and decision-making profiles in serious games analytics. Comput Human Behav 2016 Nov;64:562-574. [doi: 10.1016/j.chb.2016.07.024]
25. Wiley K, Robinson R, Mandryk RL. The making and evaluation of digital games used for the assessment of attention: systematic review. JMIR Serious Games 2021 Aug 09;9(3):e26449 [FREE Full text] [doi: 10.2196/26449] [Medline: 34383674]
26. Hagler S, Jimison HB, Pavel M. Assessing executive function using a computer game: computational modeling of cognitive processes. IEEE J Biomed Health Inform 2014 Jul;18(4):1442-1452 [FREE Full text] [doi: 10.1109/JBHI.2014.2299793] [Medline: 25014944]
27. Bang C, Nam Y, Ko EJ, Lee W, Kim B, Choi Y, et al. A serious game-derived index for detecting children with heterogeneous developmental disabilities: randomized controlled trial. JMIR Serious Games 2019 Oct 24;7(4):e14924 [FREE Full text] [doi: 10.2196/14924] [Medline: 31651408]
28. Kim HH, An JI, Park YR. A prediction model for detecting developmental disabilities in preschool-age children through digital biomarker-driven deep learning in serious games: development study. JMIR Serious Games 2021 Jun 04;9(2):e23130 [FREE Full text] [doi: 10.2196/23130] [Medline: 34085944]
29. Schoene D, Lord SR, Verhoef P, Smith ST. A novel video game--based device for measuring stepping performance and fall risk in older people. Arch Phys Med Rehabil 2011 Jul;92(6):947-953. [doi: 10.1016/j.apmr.2011.01.012] [Medline: 21549352]
30. Soancatl Aguilar V, van de Gronde JJ, Lamoth CJ, Maurits NM, Roerdink JB. Assessing dynamic balance performance during exergaming based on speed and curvature of body movements. IEEE Trans Neural Syst Rehabil Eng 2018 Jan;26(1):171-180. [doi: 10.1109/TNSRE.2017.2769701] [Medline: 29324408]
31. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: a scoping review. J Biomed Inform 2016 Dec;64:296-319 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.019] [Medline: 27815228]
32. de Klerk S, Kato PM. The future value of serious games for assessment: where do we go now? J Appl Test Technol 2017 Oct 3;18(S1):32-37.
33. Kato PM, de Klerk S. Serious games for assessment: welcome to the jungle. J Appl Test Technol 2017 Oct 3;18(S1):1-6.

XSL•FO
RenderX

34. Bamidis PD, Fissler P, Papageorgiou SG, Zilidou V, Konstantinidis EI, Billis AS, et al. Gains in cognition through combined cognitive and physical training: the role of training dosage and severity of neurocognitive disorder. Front Aging Neurosci 2015 Aug 7;7:152 [FREE Full text] [doi: 10.3389/fnagi.2015.00152] [Medline: 26300772]
35. Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD. Design, implementation, and wide pilot deployment of FitForAll: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE J Biomed Health Inform 2016 Jan;20(1):189-200. [doi: 10.1109/JBHI.2014.2378814] [Medline: 26731797]
36. Savvidis TP, Konstantinidis EI, Dias SB, Diniz JA, Hadjileontiadis LJ, Bamidis PD. Exergames for Parkinson's disease patients: how participatory design led to technology adaptation. Stud Health Technol Inform 2018;251:78-81. [Medline: 29968606]
37. Dias SB, Konstantinidis E, Diniz JA, Bamidis P, Charisis V, Hadjidimitriou S, et al. On supporting Parkinson's disease patients: the i-prognosis personalized game suite design approach. In: Proceedings of the IEEE 30th International Symposium on Computer-Based Medical Systems. 2017 Presented at: CBMS '17; June 22-24, 2017; Thessaloniki, Greece p. 521-526. [doi: 10.1109/cbms.2017.144]
38. Chrintz H, Falster O, Roed J. Single-leg postural equilibrium test. J Med Sci Sports 1991 Dec;1(4):244-246. [doi: 10.1111/j.1600-0838.1991.tb00305.x]
39. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992;83 Suppl 2:S7-11. [Medline: 1468055]
40. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
41. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. J Aging Phys Activity 1999 Apr;7(2):129-161. [doi: 10.1123/japa.7.2.129]
42. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutr Today 2015 May;50(3):117-128 [FREE Full text] [doi: 10.1097/NT.0000000000000092] [Medline: 27340299]
43. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med 1986 Mar;80(3):429-434. [doi: 10.1016/0002-9343(86)90717-5] [Medline: 3953620]
44. Graf C. The Lawton instrumental activities of daily living scale. Am J Nurs 2008 May;108(4):52-63. [doi: 10.1097/01.NAJ.0000314810.46029.74] [Medline: 18367931]
45. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: development and validation of a short-form parkinson's disease questionnaire. Psychol Health 1997 Dec;12(6):805-814. [doi: 10.1080/08870449708406741]
46. Stapleton C, Hough P, Oldmeadow L, Bull K, Hill K, Greenwood K. Four-item fall risk screening tool for subacute and residential aged care: the first step in fall prevention. Australas J Ageing 2009 Oct;28(3):139-143. [doi: 10.1111/j.1741-6612.2009.00375.x] [Medline: 19845654]
47. Mattis S. Dementia Rating Scale: Professional Manual. Odessa, FL, USA: Psychological Assessment Resources; 1988.
48. Smith A. Symbol Digit Modalities Test. Torrance, CA, USA: Western Psychological Services; 1973.
49. Υπηρεσία Φιλοξενίας Ιστοχώρων ΑΠΘ. Κέντρο Ηλεκτρονικής Διακυβέρνησης - Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης. URL: http://www.i-prognosis.eu/ [accessed 2022-09-09]
50. Chromiński K, Tkacz M. Comparison of outlier detection methods in biomedical data. J Med Informatics Technol 2010;16:89-94.
51. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978 Feb;26(1):43-49. [doi: 10.1109/TASSP.1978.1163055]
52. Dias SB, Konstantinidis E, Diniz JA, Bamidis P, Charisis V, Hadjidimitriou S, et al. Serious games as a means for holistically supporting Parkinson's Disease patients: the i-PROGNOSIS personalized game suite framework. In: Proceedings of the 9th International Conference on Virtual Worlds and Games for Serious Applications. 2017 Presented at: VS-Games '17; September 6-8, 2017; Athens, Greece p. 237-244. [doi: 10.1109/vs-games.2017.8056607]
53. Bamidis PD, Vivas AB, Styliadis C, Frantzidis C, Klados M, Schlee W, et al. A review of physical and cognitive interventions in aging. Neurosci Biobehav Rev 2014 Jul;44:206-220. [doi: 10.1016/j.neubiorev.2014.03.019] [Medline: 24705268]
54. Jimison H, Pavel M. Embedded assessment algorithms within home-based cognitive computer game exercises for elders. In: Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. 2006 Presented at: IEMBS '06; August 30-September 3, 2006; New York, NY, USA p. 6101-6104. [doi: 10.1109/iembs.2006.260303]
55. Chen X, Siebourg-Polster J, Wolf D, Czech C, Bonati U, Fischer D, et al. Feasibility of using microsoft kinect to assess upper limb movement in type III spinal muscular atrophy patients. PLoS One 2017 Jan 25;12(1):e0170472 [FREE Full text] [doi: 10.1371/journal.pone.0170472] [Medline: 28122039]
56. Gielis K, Vanden Abeele ME, Verbert K, Tournoy J, De Vos M, Vanden Abeele V. Detecting mild cognitive impairment via digital biomarkers of cognitive performance found in klondike solitaire: a machine-learning study. Digit Biomark 2021 Feb 19;5(1):44-52 [FREE Full text] [doi: 10.1159/000514105] [Medline: 33791448]
57. Gualtieri CT, Johnson LG. Neurocognitive testing supports a broader concept of mild cognitive impairment. Am J Alzheimers Dis Other Demen 2005;20(6):359-366 [FREE Full text] [doi: 10.1177/153331750502000607] [Medline: 16396441]

XSL•FO
RenderX

58. Ahmed S, de Jager C, Wilcock G. A comparison of screening tools for the assessment of mild cognitive impairment: preliminary findings. Neurocase 2012;18(4):336-351. [doi: 10.1080/13554794.2011.608365] [Medline: 22044211]
59. Shankle WR, Romney AK, Hara J, Fortier D, Dick MB, Chen JM, et al. Methods to improve the detection of mild cognitive impairment. Proc Natl Acad Sci U S A 2005 Mar 29;102(13):4919-4924 [FREE Full text] [doi: 10.1073/pnas.0501157102] [Medline: 15781874]

Abbreviations

DTW: dynamic time warping
FFT: Fullerton Fitness Test
HAC: hierarchical agglomerative clustering
SG: serious games

Edited by N Zary; submitted 08.11.21; peer-reviewed by A Teles, M Rammohan; comments to author 07.04.22; revised version received 23.06.22; accepted 21.07.22; published 13.09.22.

Please cite as:
Petsani D, Konstantinidis E, Katsouli AM, Zilidou V, Dias SB, Hadjileontiadis L, Bamidis P
Digital Biomarkers for Well-being Through Exergame Interactions: Exploratory Study
JMIR Serious Games 2022;10(3):e34768
URL: https://games.jmir.org/2022/3/e34768
doi:10.2196/34768
PMID:36099000

©Despoina Petsani, Evdokimos Konstantinidis, Aikaterini-Marina Katsouli, Vasiliki Zilidou, Sofia B Dias, Leontios Hadjileontiadis, Panagiotis Bamidis. Originally published in JMIR Serious Games (https://games.jmir.org), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effects of Acute Virtual Reality Exergaming on Mood and Executive Function: Exploratory Crossover Trial

Genta Ochi[1,2], MSc, PhD; Ryuta Kuwamizu[3], MSc; Tomomi Fujimoto[1,2], MSc, PhD; Koyuki Ikarashi[2,4], MSc; Koya Yamashiro[1,2], MSc, PhD; Daisuke Sato[1,2], MSc, PhD

[1]Department of Health and Sports, Niigata University of Health and Welfare, Niigata, Japan

[2]Institute for Human Movement and Medical Sciences, Niigata University of Health and Welfare, Niigata, Japan

[3]Laboratory of Exercise Biochemistry and Neuroendocrinology, Faculty of Health and Sport Sciences, University of Tsukuba, Tsukuba, Japan

[4]Major of Health and Welfare, Graduate School of Niigata University of Health and Welfare, Niigata, Japan

Corresponding Author:
Genta Ochi, MSc, PhD
Department of Health and Sports
Niigata University of Health and Welfare
Kita-ku, Shimami-cho, 1398
Niigata, 950-3198
Japan
Phone: 81 025 257 4595
Email: ochi@nuhw.ac.jp

Abstract

Background: Virtual reality (VR) exergaming is a new intervention strategy to help humans engage in physical activity to enhance mood. VR exergaming may improve both mood and executive function by acting on the prefrontal cortex, expanding the potential benefits. However, the impact of VR exergaming on executive function has not been fully investigated, and associated intervention strategies have not yet been established.

Objective: This study aims to investigate the effects of 10 minutes of VR exergaming on mood and executive function.

Methods: A total of 12 participants played the exergame "FitXR" under 3 conditions: (1) a VR exergame condition (ie, exercise with a head-mounted display condition [VR-EX]) in which they played using a head-mounted display, (2) playing the exergame in front of a flat display (2D-EX), and (3) a resting condition in which they sat in a chair. The color-word Stroop task (CWST), which assesses executive function; the short form of the Profile of Mood States second edition (POMS2); and the short form of the Two-Dimensional Mood Scale (TDMS), which assess mood, were administered before and after the exercise or rest conditions.

Results: The VR-EX condition increased the POMS2 vigor activity score (rest and VR-EX: t_{11}=3.69, *P*=.003) as well as the TDMS arousal (rest vs 2D-EX: t_{11}=5.34, *P*<.001; rest vs VR-EX: t_{11}=5.99, *P*<.001; 2D-EX vs VR-EX: t_{11}=3.02, *P*=.01) and vitality scores (rest vs 2D-EX: t_{11}=3.74, *P*=.007; rest vs VR-EX: t_{11}=4.84, *P*=.002; 2D-EX vs VR-EX: t_{11}=3.53, *P*=.006), suggesting that VR exergaming enhanced mood. Conversely, there was no effect on CWST performance in either the 2D-EX or VR-EX conditions. Interestingly, the VR-EX condition showed a significant positive correlation between changes in CWST arousal and reaction time (*r*=0.58, *P*=.046). This suggests that the effect of exergaming on improving executive function may disappear under an excessively increased arousal level in VR exergaming.

Conclusions: Our findings showed that 10 minutes of VR exergaming enhanced mood but did not affect executive function. This suggests that some VR content may increase cognitive demands, leading to psychological fatigue and cognitive decline as an individual approaches the limits of available attentional capacity. Future research must examine the combination of exercise and VR that enhances both brain function and mood.

(JMIR Serious Games 2022;10(3):e38200) doi:10.2196/38200

KEYWORDS

virtual reality; exergaming; exercise; executive function; physical activity; mental health

XSL•FO
RenderX

Introduction

Many previous studies support the benefits of physical activity for improving physical and mental health. Physical activity has been shown to decrease the risk of noncommunicable diseases, such as cardiovascular disease and type 2 diabetes [1], and to improve mental health [2,3]. Furthermore, in both older and younger adults, high aerobic fitness has been reported to be beneficial for maintaining executive functions (mental set shifting, information updating, and inhibition of prepotent responses [4]), making physical activity increasingly important [5-8]. However, according to the World Health Organization, approximately 25% of adults and 80% of adolescents worldwide are inactive, making physical inactivity a serious health problem. In addition, the global COVID-19 outbreak has forced people to socially disperse and self-isolate to prevent the spread of the infection, resulting in a greater proportion of individuals experiencing physical inactivity [9].

Recently, exergaming using virtual reality (VR) has been the focus of a new approach to promote physical activity [10]. The main feature of VR is that it combines a realistic 3D environment, body tracking using a head-mounted display (HMD), and handheld controllers to immerse the user in a virtual simulation [11,12]. Previous studies have suggested that VR may increase the potential for long-term participation in physical activity by distracting attention from negative images of exercise that depict it as physically fatiguing [13], boring, and strenuous [14,15], and inducing a positive mood toward exercise [16-18]. Dopaminergic nervous system involvement has been postulated as a potential brain mechanism for this positive mood effect, and VR has been used to rehabilitate patients with Parkinson disease, which is associated with decreased dopamine levels [19,20]. Previous studies have also shown that the brain's dopaminergic system, originating in the ventral tegmental area and the substantia nigra, is related to executive function through the prefrontal cortex and the striatum [21-23]. This suggests that VR exergaming may enhance executive functions governed by the prefrontal cortex, which may further augment VR's value as an exercise prescription. However, the effects on the executive function of acute exercise under VR have not been fully tested, and optimal levels of VR use remain unknown.

To examine the effects of acute exercise under VR on mood and executive function, an experimental model with high reproducibility is required to assess mood and executive function. In this study, we used the color-word Stroop task (CWST) to assess executive function, and the Two-Dimensional Mood Scale (TDMS) to assess mood state (arousal, pleasure, vitality, and stability). Previous studies have used these measures to examine exercise effects, and have found that 10 minutes of low- or moderate-intensity exercise enhanced mood states, such as arousal and pleasure, and executive function [24,25]. Furthermore, in our prior research, we identified the effects of different exercise conditions, such as listening to music [26] and a hypoxic environment [27], on mood and executive function. The 10-minute exercise set used in these studies provides an excellent model for examining VR exercises' effects on mood and executive function.

Previous studies suggested that 10 minutes of exercise enhances executive function [24,25,28-30], and that positive mood during exercise may influence the positive effect on executive function [25,26]. In this study, we hypothesized that VR is an environmental factor that enhances positive mood and executive function by exercise and aimed to clarify whether transient exercise under VR enhances both mood and executive function.

Methods

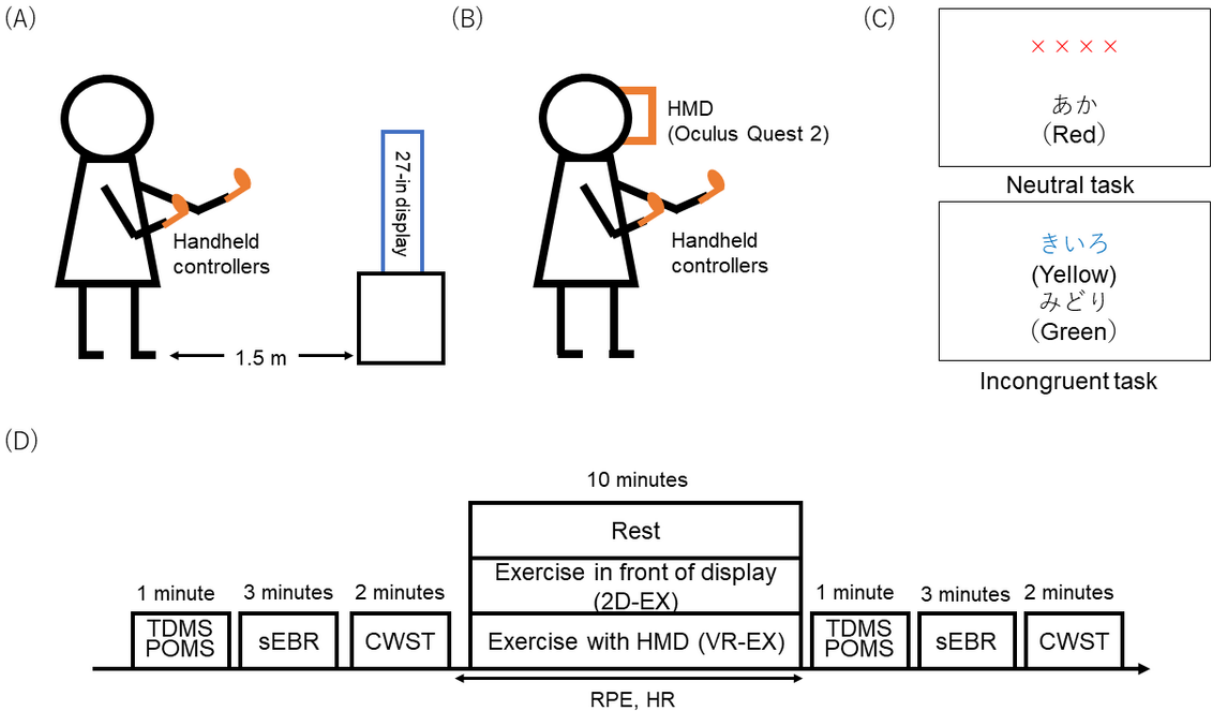
Experimental Procedure

Prior to beginning the main experimental process, each participant was given verbal instructions, guided through the consent process, practiced the CWST 3 times, and engaged in VR exergaming. Participants were instructed to play the in-game tutorial to learn how the game is played. Once the tutorial was completed, participants played the 10-minute program used in this experiment once to familiarize themselves with exergaming.

A few days after the first visit, participants engaged in 1 of the 3 experimental conditions—rest, exercise in front of a display condition (2D-EX), or exercise with an HMD (VR-EX). All participants completed all 3 conditions, each on a separate day, with the order counterbalanced across participants (Figure 1D). In the 2D-EX and VR-EX conditions, participants completed the CWST before and after 10 minutes of exercise. In the rest condition, participants completed the CWST before and after sitting in a chair for 10 minutes. In all conditions, participants completed a questionnaire and had their blink rate measured for 3 minutes before performing the CWST.

XSL•FO
RenderX

Figure 1. (A) The 2D-EX condition and (B) the VR-EX (HMD) condition. Participants performed exergaming, such as punching a target or avoiding obstacles on the display or HMD. (C) The CWST. Examples of single trials for the neutral and incongruent CWST trials presented in Japanese (English translations are in parentheses). (D) Psychological questionnaires, sEBR, and CWST were measured before and after 10 minutes of exercise or rest. 2D-EX: exercise in front of a display condition; CWST: color-word Stroop task; HMD: head-mounted display; POMS: Profile of Mood States; sEBR: spontaneous eye blink rate; TDMS: Two-Dimensional Mood Scale; VR-EX: exercise with a head-mounted display condition.



Participants

A total of 13 right-handed Japanese-speaking young adults (7 men and 6 women) were recruited from June to July 2021. The sample size was determined by assuming that the effects of exercise would be similar to those found in our previous studies [27,31]. All participants were without regular exercise habits, Japanese native speakers, healthy, and unaware of the experimental procedures for which they volunteered. No participant reported a history of neurological, psychiatric, or respiratory disorders, and none had a condition requiring medical care. One male participant withdrew from the study because he was color-blind and had difficulty discriminating colors in executive function tasks. The remaining 12 participants (6 men and 6 women) were included in the main analyses (mean age 20.15 years, SD 3.05 years); 8 participants (67%) self-reported that they played computer games on a weekly or monthly basis, and 4 participants did not specify this information. All participants were asked to refrain from exercise, alcohol consumption, and caffeine for at least 24 hours prior to each experimental session to control for outside factors that could affect cardiovascular and executive functions. Post hoc sensitivity analysis performed based on this sample with 80% power and α=.05 demonstrated sufficient sensitivity to detect repeated-measures effects exceeding f=0.50 and paired t test differences exceeding d=0.85 (with a 2-tailed α), as computed using G*Power (3.1.9.2; The G*Power Team).

Materials and Apparatus

Virtual Reality Setup

The FitXR exergame (developed by FITAR LIMITED) was administered using a commercially available HMD (Oculus Quest 2; Meta Platforms, Inc.). FitXR was selected because it is easy to play even for users with no previous exposure to VR, and can be continued for 10 minutes without interruption even if a play error occurs during the game. Using motion tracking, FitXR simulates handheld controllers as boxing gloves, whereby users punch the targets and must actively crouch and dodge obstacles. The game immerses the participant through tactile sensations of punching the target, auditory sensations of music playing, and performance feedback. In the VR-EX condition, participants played FitXR using an HMD and handheld controllers (Figure 1A).

Movie Setup for Exercise Conditions

For the same exercise condition without VR exposure (ie, the 2D-EX condition), participants exercised while watching the exergaming program movie displayed on a flat screen (27-in., 1920 × 1080 pixels, 60 Hz) positioned 1.5 m away from them (Figure 1B). This allowed participants to perform the same exercise as in the VR condition, but without the HMD. Although they performed the exercise with hand controllers in this condition, they did not receive tactile, auditory, or performance feedback.

Behavioral Measurements

The CWST, created from web-building platforms (Lab.js v19.1.0) [32], was used to evaluate executive function, and was adopted in an event-related design [8,24-30,33-35]. Two rows

XSL•FO
RenderX

containing letters or words were presented on a screen and participants were instructed to decide whether the color in which the letters or words in the top row were written corresponded to the color name presented in the bottom row (Figure 1C), pressing a "yes" or "no" button with their right or left forefinger, respectively, to respond. Reaction time (RT) and correct rate were recorded for task performance.

The CWST comprised 3 trials: 16 neutral, 16 congruent, and 16 incongruent. Each task was presented using the same method as in our previous study [27,35], and participants were asked to judge whether the color of the color name word or symbol displayed in the upper row matched the meaning of the color name word displayed in the lower row. The correct answer ratio assigned to "yes" and "no" was 50%. Each stimulus was separated by an interstimulus interval showing a fixation cross for 2 seconds to avoid a prediction of the subsequent trial's timing. The stimulus remained on the screen until a response was given or for 1 second. This study adopted Stroop interference, a specifically defined cognitive process, to elucidate the effect of an acute bout of exercise on executive function. We measured this by calculating the incongruent-neutral contrast, which is assumed to represent Stroop interference.

Physiological Measurements

Heart rate (HR) was measured with an HR sensor (H10; Polar Electro) and an HR monitor (Vantage V2; Polar Electro). Spontaneous eye blink rate (sEBR) was measured as a noninvasive brain dopaminergic system indicator [36-39]. Participants sat in front of a 27-inch display, located 70 cm from them, and were asked to look at a fixation cross presented at the center of the display at rest; the sEBR (per minute) was recorded for 3 minutes using a camcorder (120 frames/second, 2560 × 1440 pixels; Hero 8; GoPro, Inc.) that was set below the display. The sEBR was counted by 1 (GO) rater. Preliminarily, half the sample was counted by an independent rater (RK) and high validity was confirmed (r=0.995). The individual sEBR was calculated by dividing the total number of eye blinks during the 3-minute measurement interval by 3. All sEBR data were collected by 6 PM because sEBR can be less stable at night [37].

Psychological Measurements

Participant Rating of Perceived Exertion (RPE) [40] was recorded before and after the exercise intervention to assess psychological exercise intensity. In addition, the Japanese versions of the Profile of Mood States second edition (POMS2) [41] and TDMS questionnaires were administered to assess psychological indicators before and after the exercise intervention.

The POMS2 contains 35 items and evaluates 7 mood states (anger-hostility, confusion-bewilderment, depression-ejection, fatigue-inertia, tension-anxiety, vigor-activity, and friendliness). This study used a subset of 10 items related to vigor-activity and fatigue-inertia to verify whether exergaming enhances mood or causes fatigue.

The TDMS [42] is a momentary mood scale; the short form comprises 2 words describing arousal and pleasure states (lively and relaxed) and evaluates 4 mood states (arousal, pleasure, vitality, and stability). Participants were asked to indicate how they were feeling for each mood-expressing word using an 11-point Linkert scale that ranged from –5 (Listless) to 5 (Lively) and –5 (Irritated) to 5 (Relaxed). In addition to "words" and "numbers" describing the psychological state, the shortened version used "person illustrations" and "color images" to reduce the burden of answering for participants who were unfamiliar with the experiment.

Statistical Analysis

All analyses were performed using R software (4.1.2; R Foundation for Statistical Computing) and the EZR on R Commander package [43]. Two-way repeated measures ANOVA tests were performed to compare between conditions. First, the Mendoza multisample sphericity test was used to assess whether sphericity was maintained. When the sphericity assumption was confirmed, we conducted a repeated measures ANOVA with Greenhouse-Geisser epsilon correction; otherwise we performed a repeated measures ANOVA. Significant differences obtained from the ANOVA were tested using the corresponding *t* test (paired) with Holm correction. To clarify the relationships between parameters and executive performance, we conducted Pearson correlation analyses. Statistical significance was set a priori at $P<.05$ for all comparisons.

Ethics Approval

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Niigata University of Health and Welfare, Niigata, Japan (approval number 18631-210601). Participants were informed that their data would be kept confidential and provided written informed consent to participate in the study.

Results

Overview

All participants performed the experiment perfectly; no participant reported VR-related adverse effects such as motion sickness, dizziness, and headaches after VR-EX.

Physiological Parameters

Table 1 summarizes the HR, RPE, and sEBR results. These were included in the repeated measures two-way ANOVA with condition (rest/2D-EX/VR-EX) and time (before/after) as within-participant factors. The results showed a significant interaction between condition and time in HR and RPE (HR: $F_{2,22}=41.9$, $P<.001$; RPE: $F_{2,22}=39.5$, $P<.001$) and a significant main effect of time ($F_{1,11}=16.1$, $P=.002$) in sEBR. The paired *t* test with Holm correction showed significant differences in HR between the rest and 2D-EX conditions, and between the rest and VR-EX conditions during exercise (rest vs 2D-EX: $t_{11}=9.60$, $P<.001$; rest vs VR-EX: $t_{11}=7.70$, $P<.001$). The paired *t* test with Holm correction results showed significant RPE differences among all postexercise conditions (rest vs 2D-EX: $t_{11}=6.25$, $P<.001$; rest vs VR-EX: $t_{11}=7.41$, $P<.001$; 2D-EX vs VR-EX: $t_{11}=2.99$, $P=.01$).

XSL•FO
RenderX

Table 1. Physiological parameters.

Conditions and variables	Before exercise, mean (SD)	During exercise, mean (SD)	After exercise, mean (SD)
Rest			
HR[a] (bpm)	74.7 (13.7)	70.3 (13.1)	N/A[b]
RPE[c] (point)	6.0 (0.0)	N/A	6.1 (0.3)
sEBR[d] (per minute)	29.9 (20.1)	N/A	36.0 (18.5)
2D-EX[e]			
HR (bpm)	82.4 (8.8)	108.2 (15.0)[f]	N/A
RPE (point)	6.0 (0.0)	N/A	10.1 (2.3)[f]
sEBR (per minute)	30.7 (18.7)	N/A	41.2 (21.6)
VR-EX[g]			
HR (bpm)	81.0 (10.0)	117.5 (18.7)[f]	N/A
RPE (point)	6.0 (0.0)	N/A	11.7 (2.8)[f,h]
sEBR (per minute)	34.0 (21.5)	N/A	38.1 (29.2)

[a]HR: heart rate.

[b]N/A: not applicable.

[c]RPE: Rating of Perceived Exertion.

[d]sEBR: spontaneous eye blink rate.

[e]2D-EX: exercise in front of a display condition.

[f]$P<.05$ versus rest.

[g]VR-EX: exercise with a head-mounted display condition.

[h]$P<.05$ versus 2D-EX.

Psychological Parameters

The POMS2 vigor-activity and fatigue-inertia results are shown in Figure 2. The 2-way repeated measures ANOVA with condition (rest/2D-EX/VR-EX) and time (before/after) as within-participant factors showed a significant interaction between condition and time in vigor-activity ($F_{2,22}$=4.2, P=.03). The paired *t* test with Holm correction results showed significant differences between the rest and VR-EX conditions at postsession (t_{11}=3.69, P=.003).

The TDMS data in Figure 3 were analyzed using 2-way repeated measures ANOVA with condition (rest/2D-EX/VR-EX) and time (before/after) as within-participant factors, and showed a significant interaction between condition and time in arousal ($F_{1.26,13.87}$=21.14, P<.001), stability ($F_{1.31,14.36}$=12.34, P=.002), and vitality ($F_{1.42,15.67}$=18.96, P<.001), and a significant main effect of time on pleasure ($F_{1,11}$=8.99, P=.01). The paired *t* test with Holm correction results for arousal and vitality showed significant differences between the rest and 2D-EX conditions, between the rest and VR-EX conditions after exercise, and between the 2D-EX and VR-EX conditions after exercise (arousal—rest vs 2D-EX: t_{11}=5.34, P<.001; rest vs VR-EX: t_{11}=5.99, P<.001; 2D-EX vs VR-EX: t_{11}=3.02, P=.01; vitality—rest vs 2D-EX: t_{11}=3.74, P=.007; rest vs VR-EX: t_{11}=4.84, P=.002; 2D-EX vs VR-EX: t_{11}=3.53, P=.006). The paired t test with Holm correction results for stability showed significant differences between the rest and 2D-EX conditions (t_{11}=3.82, P=.006), and between the rest and VR-EX conditions (t_{11}=4.02, P=.006) after exercise.

XSL•FO
RenderX

Figure 2. (A) Vigor-activity and (B) fatigue-inertia of POMS2 under each condition. The tops and bottoms of the boxes are third and first quartiles, respectively. The upper and lower ends of the whiskers represent the highest data points within 1.5 IQRs of the upper quartiles and the lowest data points within 1.5 IQRs of the lower quartiles, respectively. The bands inside the boxes indicate medians. The red circle is the mean. *P<.05. 2D-EX: exercise in front of a display condition; POMS2: Profile of Mood States second edition; VR-EX: exercise with a head-mounted display condition.

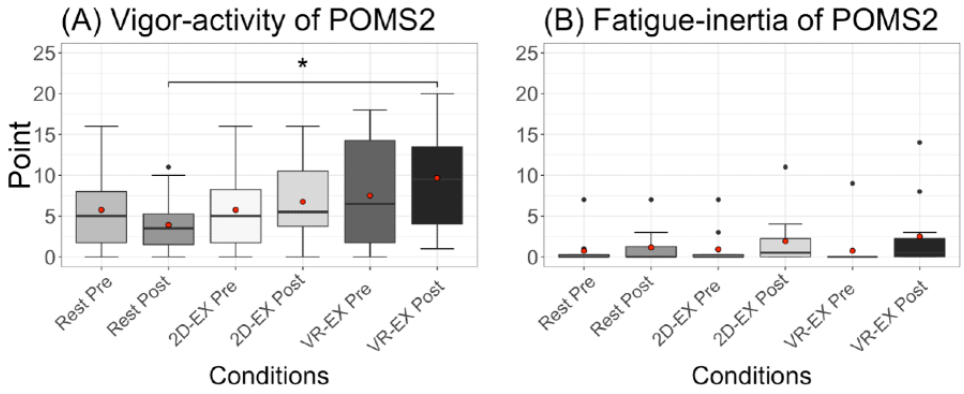
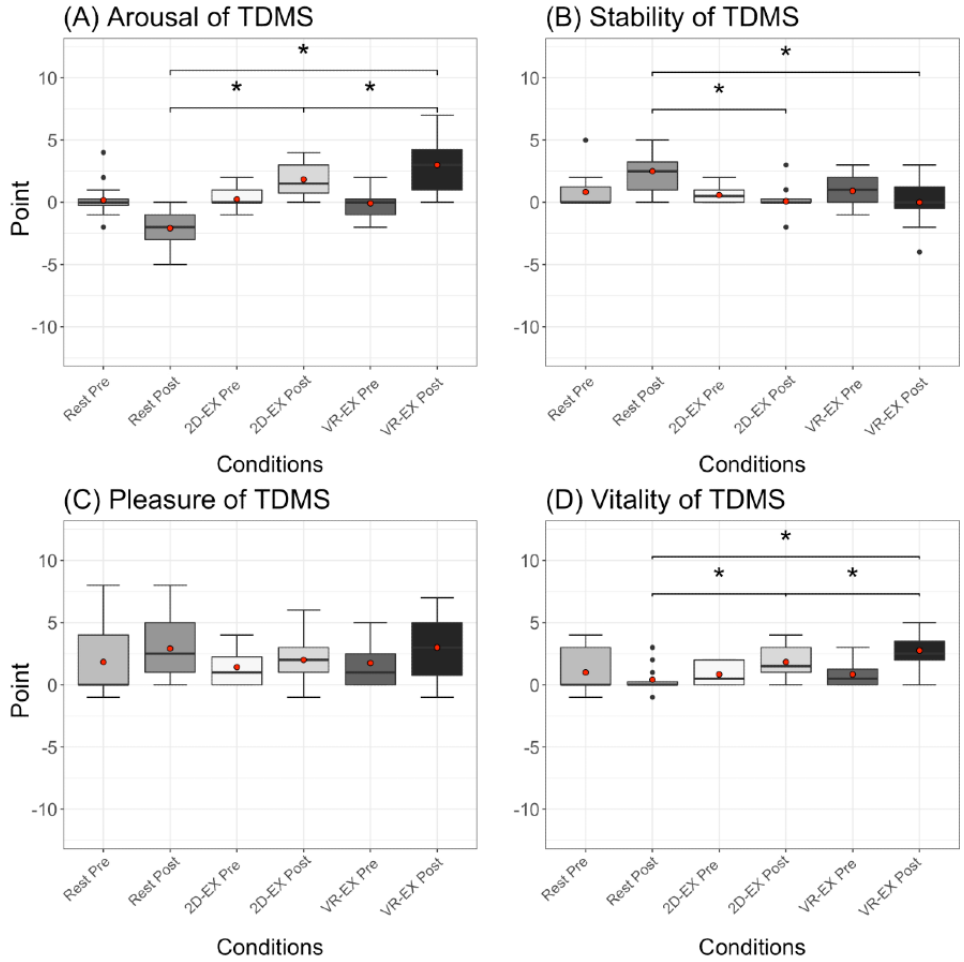


Figure 3. (A) Arousal, (B) pleasure, (C) vitality, and (D) stability of TDMS under each condition. The tops and bottoms of the boxes are third and first quartiles, respectively. The upper and lower ends of the whiskers represent the highest data points within 1.5 IQRs of the upper quartiles and the lowest data points within 1.5 IQRs of the lower quartiles, respectively. The bands inside the boxes indicate medians. The red circle is the mean. *P<.05. 2D-EX: exercise in front of a display condition; TDMS: Two-Dimensional Mood Scale; VR-EX: exercise with a head-mounted display condition.



Executive Performance: Stroop Interference

Table 2 summarizes the RT and CWST results.

First, to examine whether a general CWST tendency could be reproduced in all conditions, RT and correct rate were included in a repeated measures ANOVA using the Greenhouse-Geisser correction, with trial (neutral/incongruent), condition (rest/2D-EX/VR-EX), and time (before/after) as within-participant factors. Results showed a significant main effect of trial on RT ($F_{1.24,13.65}$=28.96, P<.001). There was no significant main effect or interaction in the correct rate ($F_{1.27,13.94}$=3.50, P=.08). This might be because the percentage of correct answers remained high, above 90% (45.83/48, 95%), for all participants. Conversely, the RT results verified that

XSL•FO
RenderX

Stroop interference was generally observed in all the sessions of this experiment. Thus, to clarify the effect of an acute bout of exercise on a specifically defined cognitive process, we analyzed Stroop interference (incongruent-neutral) by RT.

Next, to examine the RT interaction, we calculated the difference in the degree of Stroop interference between post- and presessions (incongruent-neutral presession, incongruent-neutral postsession). Then, the difference in the degree of Stroop interference between rest, 2D-EX, and VR-EX conditions was calculated using the Greenhouse-Geisser correction with condition (rest/2D-EX/VR-EX) and time (before/after) as within-participant factors. The results showed no significant main effect (condition: P=.95; time: P=.78) or interaction (P=.46; Figure 4).

Table 2. RT[a] results and CWST[b] correct rate.

Variable and conditions	Preneutral, mean (SE)	Precongruent, mean (SE)	Preincongruent, mean (SE)	Postneutral, mean (SE)	Postcongruent, mean (SE)	Postincongruent, mean (SE)
RT						
Rest	643.6 (44.6)	679.4 (42.6)	732.7 (48.4)	602.5 (27.0)	676.4 (49.4)	693.8 (45.5)
2D-EX[c]	605.4 (37.4)	637.3 (46.1)	684.3 (58.3)	607.5 (45.1)	637.6 (49.3)	698.6 (67.8)
VR-EX[d]	602.1 (31.2)	663.1 (35.3)	666.5 (45.0)	597.7 (49.4)	615.0 (40.2)	658.8 (50.4)
Correct rate						
Rest	96.8 (1.2)	96.4 (1.2)	93.2 (2.0)	96.9 (1.2)	96.9 (1.4)	92.7 (2.2)
2D-EX	97.9 (1.2)	95.3 (1.7)	93.2 (2.1)	96.4 (1.4)	95.3 (1.7)	97.4 (1.6)
VR-EX	99.0 (0.7)	96.4 (0.9)	93.2 (2.1)	96.4 (1.2)	97.4 (1.2)	96.9 (1.6)

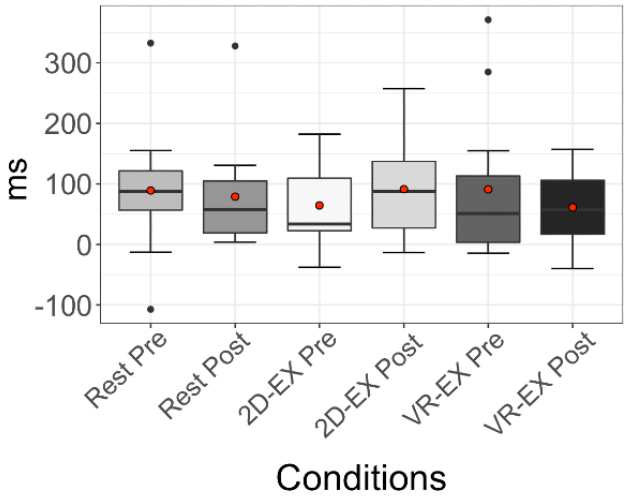
[a]RT: reaction time.

[b]CWST: color-word Stroop task.

[c]2D-EX: exercise in front of a display condition.

[d]VR-EX: exercise with a head-mounted display condition.

Figure 4. Stroop interference (RT) differences between pre- and postsessions for each condition. Stroop interference differences showed no significant main effect or interaction. The tops and bottoms of the boxes are third and first quartiles, respectively. The upper and lower ends of the whiskers represent the highest data points within 1.5 IQRs of the upper quartiles and the lowest data points within 1.5 IQRs of the lower quartiles, respectively. The bands inside the boxes indicate the medians. The red circle is the mean. 2D-EX: exercise in front of a display condition; RT: reaction time; VR-EX: exercise with a head-mounted display condition.

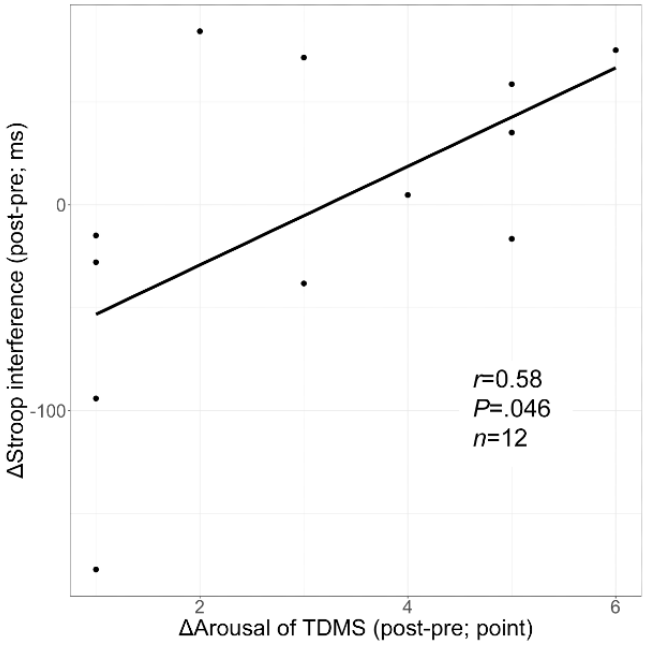


Association Between Executive Performance and Physiological and Psychological Results

We examined the correlation between the change in Stroop interference–related RT and altered physiological and psychological parameters in the 2D-EX and VR-EX conditions. The VR-EX condition showed a significant correlation between Stroop interference and arousal (r=0.58, P=.046; Figure 5). In the 2D-EX condition, there was no correlation between Stroop interference and any parameters. Furthermore, neither condition showed a correlation between the exercise-induced change in Stroop interference and sEBR before exercise.

XSL•FO
RenderX

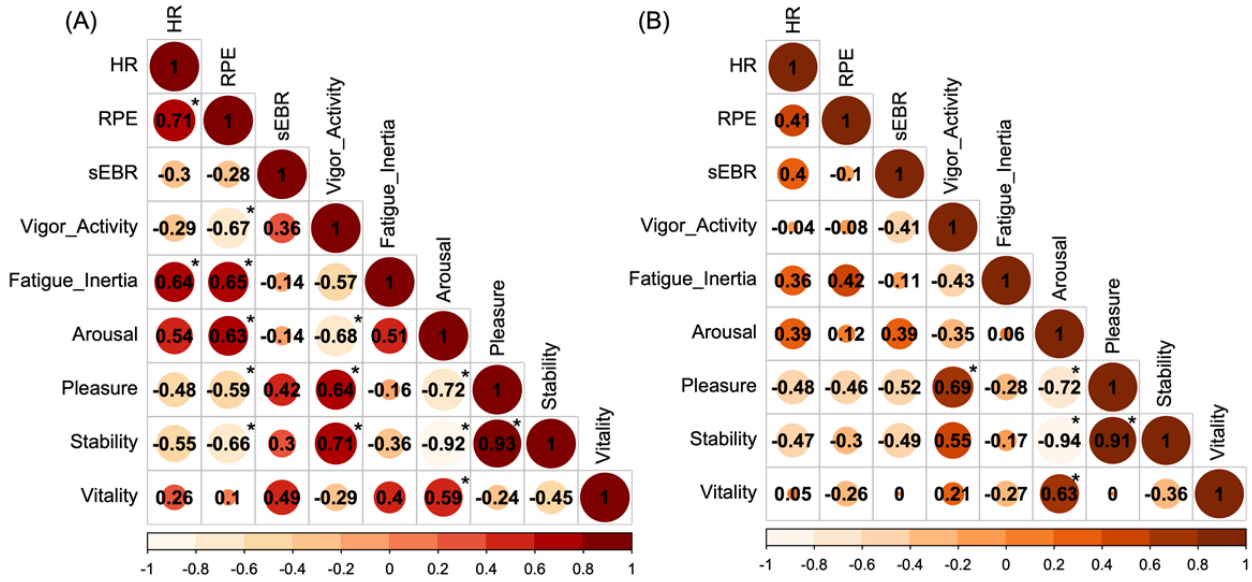
Figure 5. The association between Δarousal (post-pre) and ΔStroop interference (post-pre) in the VR-EX condition. TDMS: Two-Dimensional Mood Scale; VR-EX: exercise with a head-mounted display condition.



Association Between Physiological and Psychological Results

We examined the correlation between physiological and psychological parameters in the 2D-EX and VR-EX conditions. The correlation matrix summarizing the correlation coefficients and significance results between each variable in the VR-EX and 2D-EX conditions is shown in Figure 6. The VR-EX condition showed correlations between HR and fatigue-inertia (r=0.64, P=.02), RPE and vigor-activity (r=–0.67, P=.02), fatigue-inertia (r=0.65, P=.02), arousal (r=0.63, P=.02), pleasure (r=–0.59, P=.04), and stability (r=–0.66, P=.02). The 2D-EX condition showed no correlations between physiological and psychological parameters. sEBR was not related to any of the psychological parameters in either condition.

Figure 6. The correlation matrix of physiological and psychological parameters in the (A) VR-EX and (B) 2D-EX conditions. The numbers in the matrix indicate the correlation coefficients between each variable, with the asterisk indicating significant correlations (P<.05). Color intensity indicates positive and negative correlation coefficients, and the size of the circle indicates the strength of r. 2D-EX: exercise in front of a display condition; HR: heart rate; RPE: Rating of Perceived Exertion; sEBR: spontaneous eye blink rate; VR-EX: exercise with a head-mounted display condition.



XSL•FO
RenderX

Discussion

Principal Findings

This study investigated whether VR exergaming enhances executive function and mood. The findings showed that VR exergaming enhanced positive moods, such as vigor-activity and vitality, but did not improve executive function. These results might be explained by the increased arousal level after exercise.

Using estimates of traditional predicted maximum HR [44], we calculated the percent maximum HRs during exercise as 54.1% (2D-EX) and 58.8% (VR-EX). In addition, RPE, which measures a psychological evaluation of exercise intensity, increased to the same extent as in previous studies using a 10-minute moderate-intensity exercise intervention [26,34]. These results suggest that the 2D-EX and VR-EX conditions in this study could be considered as moderate-intensity exercises [45]. As there was no difference in HR between conditions, we believe that the conditions induced the same exercise intensity.

The VR-EX condition caused an increase in the POMS2 vigor-activity scale. Both the 2D-EX and VR-EX conditions increased TDMS arousal, stability, and vitality levels, but arousal and vitality levels differed between the 2 conditions. Participants performed the same-intensity exercise in both conditions, suggesting that the VR exposure might have synergistically increased arousal and vitality levels. These results replicate previous findings that VR exercise improves mood [16]. Conversely, no change in pleasure level was observed in either the 2D-EX or VR-EX condition. As the VR-EX condition involves exergaming to music, we hypothesized that the pleasure level would increase after exercise, because previous studies found that the combination of music and exercise increased the pleasure level [26]; however, this hypothesis was not supported. Although no study participants exhibited symptoms or self-reported VR sickness, it is possible that discomfort from the HMD affected their pleasure level [46,47]. These results might demonstrate the unique characteristics of VR as an exercise environment.

We checked whether Stroop interference, which assesses executive function, was occurring in this experiment. The behavioral measurements revealed a shorter RT in the neutral trials than in the incongruent trials. Thus, we confirmed that Stroop interference could be induced before and after an acute bout of exercise or rest in all conditions. Based on these results, we first compared the effect of all conditions on Stroop interference and found no significant change between conditions; besides, there was no improvement in executive function in either the 2D-EX or VR-EX condition. In addition, there were no differences in executive function improvement between the VR-EX and 2D-EX conditions, despite differences in vigor, arousal, and vitality levels. These results were contrary to our hypothesis and differed from previous studies showing that 10 minutes of moderate-intensity exercise enhanced executive function [30]. Therefore, we examined the physiological and psychological indicators that influenced the changes in executive function in this study.

Our analysis showed an increase in arousal levels as measured by the TDMS in both the 2D-EX and VR-EX conditions, but contrary to previous studies, no association was found in the 2D-EX condition and the opposite association was found in the VR-EX condition. This phenomenon may be considered from the inverse U-shaped model of arousal levels [48]. VR has been shown to increase emotional arousal recorded by multichannel electroencephalograms [49,50], which is consistent with our results showing that the VR-EX condition had a greater arousal increase than the 2D-EX condition. Although exercise-induced increases in intraparticipant arousal levels measured by the TDMS have been associated with enhanced cognitive function and prefrontal activity, which controls cognitive processing [25], excessively increased arousal levels related to high intensity exercise [51] or stress [52] may counteract the beneficial effects of exercise on cognitive function. Therefore, the combination of exercise and VR in this study might have excessively increased arousal levels, counteracting the beneficial effect of exercise on executive function.

Improved executive function from exercise is not only related to arousal level but also to pleasure level [26]. The brain dopaminergic system is related to executive function through prefrontal cortex functions [21-23]. We hypothesized that VR increased the pleasure level and executive function by activating the brain dopaminergic system during exercise. However, our findings showed no change in pleasure levels measured by the TDMS and sEBR (a noninvasive brain dopaminergic system indicator) in the VR-EX condition. This may be because of participants' lack of familiarity with the exercise style and VR exergaming. No study participants had prior experience with VR exergaming. In addition, many previous studies in which exercise under VR elicited positive mood used bicycle exercises [16-18]; however, participants are more likely to be familiar with a bicycling exercise style than the boxing motion exercise used in this study. Future studies should thus examine long-term intervention effects to familiarize participants with VR and exercise style prior to the VR exergaming intervention.

Furthermore, our findings showed no improvement in executive function in either the VR-EX or 2D-EX condition. The 2D-EX condition in this study combined exercise and computer games, where participants exercised according to the target shown on the display. Therefore, our results are consistent with those of previous studies reporting that a combination of transient exercise and games did not improve executive function [53,54]. Dual tasking has been shown to cause psychological fatigue and cognitive decline (cognitive fatigue) as physical and mental cognitive demands increase [54-57]. However, previous studies in which transient exercise improved executive function have involved only simple bicycle or running exercises [25,26,28,30]. Therefore, the 2D-EX and VR-EX conditions might have increased cognitive demands during exercise, causing cognitive fatigue, which may have counteracted the effect of the exercise on improving executive function. Further research is needed to more comprehensively investigate exercise combined with VR; it is also necessary to examine an exercise-alone condition.

XSL•FO
RenderX

Limitations

Several limitations of this study should be noted. First, it included only healthy adults without regular exercise habits; thus, there was insufficient interparticipant evaluation. Further validation with athletes, children, and the elderly is needed to clarify the effectiveness of VR exergaming. Second, the TDMS, which is a psychological rather than a physiological scale, was used to assess brain arousal levels. Future studies should assess physiological arousal using electroencephalograms to clarify optimal VR exergaming conditions that enhance brain function based on brain arousal level. Third, this study used FitXR, a VR exergame with a set exercise intensity; it is unclear to what extent the study results generalize to other games. Our findings suggest that games with high cognitive demands might eliminate the positive effect of exercise on executive function; however, further research is needed to determine whether games with low cognitive demands improve executive function. In addition, as mentioned earlier, we examined the transient intervention effects of VR exergaming; however, long-term intervention effects remain unclear. VR exergaming can cause negative effects, such as VR sickness [46], and positive effects, such as improved mood. It is thus necessary to identify the effects of long-term VR exergaming interventions on the mind and body to develop a strategy for promoting physical activity.

Conclusions

Our study confirms that VR exergaming improves some mood items, but did not provide evidence that it improves executive function. We found that VR exergaming with high cognitive demands may inhibit the positive effects of exercise on executive function. To propose VR exergaming that promotes exercise habituation, it is essential to further examine the combination of exercise and VR conditions that enhances both brain function and mood.

Acknowledgments

GO worked in the conceptualization of this research, played a role in data collection and analysis, and wrote the manuscript. DS and RK completed data analysis and discussion. KI, TF, and KY assisted in editing the final drafts of the manuscript. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grant JP19K20036 (GO). We thank Editage for English language editing.

Conflicts of Interest

None declared.

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet 2012 Jul;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9]
2. Blondell SJ, Hammersley-Mather R, Veerman JL. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. BMC Public Health 2014 May 27;14:510 [FREE Full text] [doi: 10.1186/1471-2458-14-510] [Medline: 24885250]
3. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
4. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol 2000 Aug;41(1):49-100. [doi: 10.1006/cogp.1999.0734] [Medline: 10945922]
5. Hyodo K, Dan I, Kyutoku Y, Suwabe K, Byun K, Ochi G, et al. The association between aerobic fitness and cognitive function in older men mediated by frontal lateralization. Neuroimage 2016 Jan 15;125:291-300. [doi: 10.1016/j.neuroimage.2015.09.062] [Medline: 26439424]
6. Themanson JR, Pontifex MB, Hillman CH. Fitness and action monitoring: evidence for improved cognitive flexibility in young adults. Neuroscience 2008 Nov 19;157(2):319-328 [FREE Full text] [doi: 10.1016/j.neuroscience.2008.09.014] [Medline: 18845227]
7. Weinstein AM, Voss MW, Prakash RS, Chaddock L, Szabo A, White SM, et al. The association between aerobic fitness and executive function is mediated by prefrontal cortex volume. Brain Behav Immun 2012 Jul;26(5):811-819 [FREE Full text] [doi: 10.1016/j.bbi.2011.11.008] [Medline: 22172477]
8. Kuwamizu R, Suwabe K, Damrongthai C, Fukuie T, Ochi G, Hyodo K, et al. Spontaneous Eye Blink Rate Connects Missing Link between Aerobic Fitness and Cognition. Med Sci Sports Exerc 2021 Jul 01;53(7):1425-1433. [doi: 10.1249/MSS.0000000000002590] [Medline: 33433152]
9. Ammar A, Trabelsi K, Brach M, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of home confinement on mental health and lifestyle behaviours during the COVID-19 outbreak: insights from the ECLB-COVID19 multicentre study. Biol Sport 2021 Mar;38(1):9-21 [FREE Full text] [doi: 10.5114/biolsport.2020.96857] [Medline: 33795912]

XSL•FO
RenderX

10. Ahn SJ, Fox J. Immersive Virtual Environments, Avatars, and Agents for Health: Oxford Research Encyclopedias Communication. ScienceGate. 2017. URL: https://www.sciencegate.app/document/10.1093/acrefore/9780190228613.013.325 [accessed 2022-09-19]
11. Witmer BG, Singer MJ. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
12. Slater M, Wilbur S. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. Presence: Teleoperators & Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
13. Faric N, Potts HWW, Hon A, Smith L, Newby K, Steptoe A, et al. What Players of Virtual Reality Exercise Games Want: Thematic Analysis of Web-Based Reviews. J Med Internet Res 2019 Sep 16;21(9):e13833 [FREE Full text] [doi: 10.2196/13833] [Medline: 31538951]
14. Qian J, McDonough DJ, Gao Z. The Effectiveness of Virtual Reality Exercise on Individual's Physiological, Psychological and Rehabilitative Outcomes: A Systematic Review. Int J Environ Res Public Health 2020 Jun 10;17(11):4133 [FREE Full text] [doi: 10.3390/ijerph17114133] [Medline: 32531906]
15. Ekkekakis P, Hall EE, Petruzzello SJ. The relationship between exercise intensity and affective responses demystified: to crack the 40-year-old nut, replace the 40-year-old nutcracker!. Ann Behav Med 2008 Apr;35(2):136-149. [doi: 10.1007/s12160-008-9025-z] [Medline: 18369689]
16. Zeng N, Pope Z, Gao Z. Acute Effect of Virtual Reality Exercise Bike Games on College Students' Physiological and Psychological Outcomes. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):453-457. [doi: 10.1089/cyber.2017.0042] [Medline: 28715263]
17. Plante TG, Aldridge A, Bogden R, Hanelin C. Might virtual reality promote the mood benefits of exercise? Computers in Human Behavior 2003 Jul;19(4):495-509. [doi: 10.1016/S0747-5632(02)00074-2]
18. Mestre DR, Dagonneau V, Mercier C. Does Virtual Reality Enhance Exercise Performance, Enjoyment, and Dissociation? An Exploratory Study on a Stationary Bike Apparatus. Presence: Teleoperators and Virtual Environments 2011 Feb 01;20(1):1-14. [doi: 10.1162/pres_a_00031]
19. Triegaardt J, Han TS, Sada C, Sharma S, Sharma P. The role of virtual reality on outcomes in rehabilitation of Parkinson's disease: meta-analysis and systematic review in 1031 participants. Neurol Sci 2020 Mar;41(3):529-536 [FREE Full text] [doi: 10.1007/s10072-019-04144-3] [Medline: 31808000]
20. Juras G, Brachman A, Michalska J, Kamieniarz A, Pawłowski M, Hadamus A, et al. Standards of Virtual Reality Application in Balance Training Programs in Clinical Practice: A Systematic Review. Games Health J 2019 Apr;8(2):101-111. [doi: 10.1089/g4h.2018.0034] [Medline: 30239217]
21. Vernaleken I, Buchholz H, Kumakura Y, Siessmeier T, Stoeter P, Bartenstein P, et al. 'Prefrontal' cognitive performance of healthy subjects positively correlates with cerebral FDOPA influx: an exploratory [18F]-fluoro-L-DOPA-PET investigation. Hum Brain Mapp 2007 Oct;28(10):931-939 [FREE Full text] [doi: 10.1002/hbm.20325] [Medline: 17133402]
22. Nagano-Saito A, Leyton M, Monchi O, Goldberg YK, He Y, Dagher A. Dopamine depletion impairs frontostriatal functional connectivity during a set-shifting task. J Neurosci 2008 Apr 02;28(14):3697-3706 [FREE Full text] [doi: 10.1523/JNEUROSCI.3921-07.2008] [Medline: 18385328]
23. Cools R. Chemistry of the Adaptive Mind: Lessons from Dopamine. Neuron 2019 Oct 09;104(1):113-131 [FREE Full text] [doi: 10.1016/j.neuron.2019.09.035] [Medline: 31600509]
24. Damrongthai C, Kuwamizu R, Suwabe K, Ochi G, Yamazaki Y, Fukuie T, et al. Benefit of human moderate running boosting mood and executive function coinciding with bilateral prefrontal activation. Sci Rep 2021 Nov 22;11(1):22657 [FREE Full text] [doi: 10.1038/s41598-021-01654-z] [Medline: 34811374]
25. Byun K, Hyodo K, Suwabe K, Ochi G, Sakairi Y, Kato M, et al. Positive effect of acute mild exercise on executive function via arousal-related prefrontal activations: an fNIRS study. Neuroimage 2014 Sep;98:336-345. [doi: 10.1016/j.neuroimage.2014.04.067] [Medline: 24799137]
26. Suwabe K, Hyodo K, Fukuie T, Ochi G, Inagaki K, Sakairi Y, et al. Positive Mood while Exercising Influences Beneficial Effects of Exercise with Music on Prefrontal Executive Function: A Functional NIRS Study. Neuroscience 2021 Feb 01;454:61-71. [doi: 10.1016/j.neuroscience.2020.06.007] [Medline: 32554109]
27. Ochi G, Yamada Y, Hyodo K, Suwabe K, Fukuie T, Byun K, et al. Neural basis for reduced executive performance with hypoxic exercise. Neuroimage 2018 May 01;171:75-83. [doi: 10.1016/j.neuroimage.2017.12.091] [Medline: 29305162]
28. Kujach S, Byun K, Hyodo K, Suwabe K, Fukuie T, Laskowski R, et al. A transferable high-intensity intermittent exercise improves executive performance in association with dorsolateral prefrontal activation in young adults. Neuroimage 2018 Apr 01;169:117-125. [doi: 10.1016/j.neuroimage.2017.12.003] [Medline: 29203453]
29. Hyodo K, Dan I, Suwabe K, Kyutoku Y, Yamada Y, Akahori M, et al. Acute moderate exercise enhances compensatory brain activation in older adults. Neurobiol Aging 2012 Nov;33(11):2621-2632. [doi: 10.1016/j.neurobiolaging.2011.12.022] [Medline: 22300952]
30. Yanagisawa H, Dan I, Tsuzuki D, Kato M, Okamoto M, Kyutoku Y, et al. Acute moderate exercise elicits increased dorsolateral prefrontal activation and improves cognitive performance with Stroop test. Neuroimage 2010 May 01;50(4):1702-1710. [doi: 10.1016/j.neuroimage.2009.12.023] [Medline: 20006719]

XSL•FO
RenderX

31. Ochi G, Kuwamizu R, Suwabe K, Fukuie T, Hyodo K, Soya H. Cognitive fatigue due to exercise under normobaric hypoxia is related to hypoxemia during exercise. Sci Rep 2022 Jun 28;12(1):9835 [FREE Full text] [doi: 10.1038/s41598-022-14146-5] [Medline: 35764684]
32. Shevchenko Y. Open Lab: A web application for running and sharing online experiments. Behav Res Methods 2022 Mar 01;1:1-1. [doi: 10.3758/s13428-021-01776-2] [Medline: 35233751]
33. Zysset S, Müller K, Lohmann G, von Cramon DY. Color-word matching stroop task: separating interference and response conflict. Neuroimage 2001 Jan;13(1):29-36. [doi: 10.1006/nimg.2000.0665] [Medline: 11133306]
34. Suwabe K, Hyodo K, Byun K, Ochi G, Yassa MA, Soya H. Acute moderate exercise improves mnemonic discrimination in young adults. Hippocampus 2017 Mar;27(3):229-234 [FREE Full text] [doi: 10.1002/hipo.22695] [Medline: 27997992]
35. Ochi G, Kanazawa Y, Hyodo K, Suwabe K, Shimizu T, Fukuie T, et al. Hypoxia-induced lowered executive function depends on arterial oxygen desaturation. J Physiol Sci 2018 Nov;68(6):847-853. [doi: 10.1007/s12576-018-0603-y] [Medline: 29536370]
36. Eckstein MK, Guerra-Carrillo B, Miller Singley AT, Bunge SA. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? Dev Cogn Neurosci 2017 Jun;25:69-91 [FREE Full text] [doi: 10.1016/j.dcn.2016.11.001] [Medline: 27908561]
37. Jongkees BJ, Colzato LS. Spontaneous eye blink rate as predictor of dopamine-related cognitive function-A review. Neurosci Biobehav Rev 2016 Dec;71:58-82. [doi: 10.1016/j.neubiorev.2016.08.020] [Medline: 27555290]
38. Groman SM, James AS, Seu E, Tran S, Clark TA, Harpster SN, et al. In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. J Neurosci 2014 Oct 22;34(43):14443-14454 [FREE Full text] [doi: 10.1523/JNEUROSCI.3037-14.2014] [Medline: 25339755]
39. Karson CN. Spontaneous eye-blink rates and dopaminergic systems. Brain 1983 Sep;106 (Pt 3):643-653. [doi: 10.1093/brain/106.3.643] [Medline: 6640274]
40. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
41. Heuchert PJ, McNair DM. Profile of Mood States 2nd Edition Japanese Manual (in Japanese). Tokyo: Kaneko Shobo; Sep 19, 2017.
42. Sakairi Y, Nakatsuka K, Shimizu T. Development of the Two-Dimensional Mood Scale for self-monitoring and self-regulation of momentary mood states. Jpn Psychol Res 2013 Jun 05;55(4):338-349. [doi: 10.1111/jpr.12021]
43. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013 Mar;48(3):452-458 [FREE Full text] [doi: 10.1038/bmt.2012.244] [Medline: 23208313]
44. Fox SM, Naughton JP. Physical activity and the prevention of coronary heart disease. Prev Med 1972 Mar;1(1):92-120. [doi: 10.1016/0091-7435(72)90079-5] [Medline: 5069016]
45. American College of Sports Medicine (ACSM). ACSM's guidelines for exercise testing and prescription. Alphen aan den Rijn, The Netherlands: Wolters Kluwer Health; 2014.
46. Szpak A, Michalski SC, Loetscher T. Exergaming With Beat Saber: An Investigation of Virtual Reality Aftereffects. J Med Internet Res 2020 Oct 23;22(10):e19840 [FREE Full text] [doi: 10.2196/19840] [Medline: 33095182]
47. Simone LK, Schultheis MT, Rebimbas J, Millis SR. Head-mounted displays for clinical virtual reality applications: pitfalls in understanding user behavior while using technology. Cyberpsychol Behav 2006 Oct;9(5):591-602. [doi: 10.1089/cpb.2006.9.591] [Medline: 17034327]
48. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol 1908 Nov;18(5):459-482. [doi: 10.1002/cne.920180503]
49. Ding N, Zhou W, Fung AY. Emotional effect of cinematic VR compared with traditional 2D film. Telematics and Informatics 2018 Sep;35(6):1572-1579. [doi: 10.1016/j.tele.2018.04.003]
50. Tian F, Hua M, Zhang W, Li Y, Yang X. Emotional arousal in 2D versus 3D virtual reality environments. PLoS One 2021;16(9):e0256211 [FREE Full text] [doi: 10.1371/journal.pone.0256211] [Medline: 34499667]
51. Kamijo K, Nishihira Y, Hatta A, Kaneda T, Wasaka T, Kida T, et al. Differential influences of exercise intensity on information processing in the central nervous system. Eur J Appl Physiol 2004 Jul;92(3):305-311. [doi: 10.1007/s00421-004-1097-2] [Medline: 15083372]
52. Joseph R. The neurology of traumatic "dissociative" amnesia: commentary and literature review. Child Abuse Negl 1999 Aug;23(8):715-727. [doi: 10.1016/s0145-2134(99)00048-4] [Medline: 10477233]
53. Kamijo K, Abe R. Aftereffects of Cognitively Demanding Acute Aerobic Exercise on Working Memory. Med Sci Sports Exerc 2019 Jan;51(1):153-159 [FREE Full text] [doi: 10.1249/MSS.0000000000001763] [Medline: 30153193]
54. Douris PC, Handrakis JP, Apergis D, Mangus RB, Patel R, Limtao J, et al. The Effects of Aerobic Exercise and Gaming on Cognitive Performance. J Hum Kinet 2018 Mar;61:73-83 [FREE Full text] [doi: 10.1515/hukin-2017-0134] [Medline: 29599861]
55. McDowd JM. An overview of attention: behavior and brain. J Neurol Phys Ther 2007 Sep;31(3):98-103. [doi: 10.1097/NPT.0b013e31814d7874] [Medline: 18025955]
56. Marois R, Ivanoff J. Capacity limits of information processing in the brain. Trends Cogn Sci 2005 Jun;9(6):296-305. [doi: 10.1016/j.tics.2005.04.010] [Medline: 15925809]

XSL•FO
RenderX

57. Borragán G, Slama H, Bartolomei M, Peigneux P. Cognitive fatigue: A Time-based Resource-sharing account. Cortex 2017 Apr;89:71-84. [doi: 10.1016/j.cortex.2017.01.023] [Medline: 28237888]

Abbreviations

2D-EX: exercise in front of a display condition
CWST: color-word Stroop task
HMD: head-mounted display
HR: heart rate
POMS2: Profile of Mood States second edition
RPE: Rating of Perceived Exertion
RT: reaction time
sEBR: spontaneous eye blink rate
TDMS: Two-Dimensional Mood Scale
VR: virtual reality
VR-EX: exercise with a head-mounted display condition

Edited by N Zary; submitted 23.03.22; peer-reviewed by W Deng, E Sadeghi-Demneh; comments to author 22.07.22; revised version received 03.08.22; accepted 02.09.22; published 28.09.22.

Please cite as:
Ochi G, Kuwamizu R, Fujimoto T, Ikarashi K, Yamashiro K, Sato D
The Effects of Acute Virtual Reality Exergaming on Mood and Executive Function: Exploratory Crossover Trial
JMIR Serious Games 2022;10(3):e38200
URL: https://games.jmir.org/2022/3/e38200
doi:10.2196/38200
PMID:36169992

©Genta Ochi, Ryuta Kuwamizu, Tomomi Fujimoto, Koyuki Ikarashi, Koya Yamashiro, Daisuke Sato. Originally published in JMIR Serious Games (https://games.jmir.org), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Cognitive Function on Virtual Reality Intervention for Upper Extremity Rehabilitation of Patients With Subacute Stroke: Prospective Randomized Controlled Trial With 6-Month Follow-up

Yan Leng[1*], PhD; Wai Leung Ambrose Lo[1,2*], PhD; Yu Rong Mao[2,3], MSc; Ruihao Bian[1], BSc; Jiang Li Zhao[1], MSc; Zhiqin Xu[1], BSc; Le Li[4], PhD; Dong Feng Huang[2,3], MD

[1]Department of Rehabilitation Medicine, First Affiliated Hospital Sun Yat-sen University, Guangzhou, China

[2]Guangdong Engineering and Technology Research Center for Rehabilitation Medicine and Translation, Sun Yat-sen University, Guangzhou, China

[3]Department of Rehabilitation Medicine, Seventh Affiliated Hospital Sun Yat-sen University, Shenzhen, China

[4]Institute of Medical Research, Northwestern Polytechnical University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Dong Feng Huang, MD
Guangdong Engineering and Technology Research Center for Rehabilitation Medicine and Translation
Sun Yat-sen University
58 Zhongshan Er Lu Road
Guangzhou, 510080
China
Phone: 86 13922318919
Email: huangdf@mail.sysu.edu.cn

Abstract

Background: Stroke is among the leading causes of long-term disability worldwide. Motor impairments after stroke not only impact the individuals quality of life but also lay substantial burdens on the society. Motor planning is a key component of cognitive function that impacts motor control. Hand movements such as grasping or reaching to grasp require the application of correct force and the coordination of multiple limb segments. Successful completion of hand motor task requires a certain degree of cognitive function to anticipate the requirement of the task. Cognitive function may thus be a confounding factor to rehabilitation outcomes.

Objective: This study aims to explore the impact of cognitive function on functional outcomes in people with subacute stroke after VR intervention.

Methods: Patients with stroke were first stratified into cognitively normal (CN) and cognitively impaired (CI), followed by allocation to the VR or control group (CG). Fugl-Meyer Assessment for Upper Extremity (FMA-UE), Barthel Index (BI), and Instrumental Activities of Daily Living (IADL) were recorded at baseline, 3 weeks after the intervention, and 3 and 6 months after the intervention. The between-group and within-group differences were assessed by repeated-measures analysis of variance (ANOVA).

Results: The between-group comparison indicated that FMA-UE, BI, and IADL (time effect $P<.001$ for all) scores improved significantly in both groups after the intervention. Repeated-measures ANOVA indicated that FMA-UE, BI, and IADL (time effect $P<.001$ for all) were significantly different in each subgroup after the intervention. For BI score, the ANOVA results showed obvious interaction effects (treatment × time × cognitive effect, $P=.04$).

Conclusions: VR intervention was as effective as traditional conventional therapy in improving upper limb function regardless of the cognitive functional level. Patients with stroke with impaired cognitive function may gain more improvement in upper limb function and independency in performing activities of daily living after a VR-based intervention.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOC-15006064; https://tinyurl.com/4c9vkrrn

(JMIR Serious Games 2022;10(3):e33755) doi:10.2196/33755

XSL•FO
RenderX

KEYWORDS

stroke; motor function of upper extremity; virtual reality; cognitive function

Introduction

Stroke is among the leading causes of long-term disability worldwide [1]. Despite the continuous improvement in rehabilitation technology, approximately 80% of acute and subacute stroke survivors continue to have residual upper extremity dysfunction of varying degrees [2]. Motor impairments after stroke not only impact the individuals quality of life but also lay substantial burdens on the society.

Virtual reality (VR) intervention is considered a promising approach in stroke rehabilitation. It is characterized by task-oriented and repetitive training with cognitive training elements [3,4]. VR systems create a simulated real life or imaginary environment where participants could interact dynamically [3,5]. Most published studies indicated that this technology provides a variable rehabilitation approach that improves physical function and reduces the demand on staff time [3]. A functional magnetic resonance image (fMRI) study previously published indicated cortical reorganization after VR training, which corresponded with upper limb motor function improvement [6]. Other studies also reported that the VR game system is superior to conventional therapy alone in improving upper extremity motor function recovery when used in conjunction with other interventions [3,7,8]. However, conflicting evidence was reported that VR intervention did not result in better function of the upper extremity when compared with traditional therapy [9,10], or that the change of the motor function was similar between the VR group and the control group [11,12]. A potential reason for the conflicting results may be related to the cognitive status of participants at baseline level.

Motor planning is a key component of cognitive function that impacts motor control [13]. It is known that cognitive function is a predictor of functional outcome in people with stroke [14]. Hand movements such as grasping or reaching to grasp require the application of correct force and the coordination of multiple limb segments [15]. Successful completion of hand motor task must therefore require a certain degree of cognitive function to anticipate the requirement of the task [16,17]. Thus, cognitive function may be a confounding factor of rehabilitation outcome. Studies that employed electro capnography to investigate the cognitive neural process of motor planning reported an increase in computational demand from bilateral hemispheres in patients with stroke [18,19]. A published study that utilized transcranial magnetic stimulation did not show direct evidence that the improvement in cortical activity is clinically relevant to upper limb function [20]. A potential reason is that transcranial magnetic stimulation (TMS) intervention has minimal cognitive involvement, which may limit its benefit on motor skills recovery. Initial studies in this area indicated that patients with higher cognitive status at the time of admission tended to have better rehabilitation outcomes [21,22]. By contrast, some other studies concluded that cognitive impairment had no negative effect on functional improvement [9,23]. Diamond et al [24] also proposed that the reason for poor functional outcome among patients with stroke having cognitive impairment may be more related to the low motor functional status at the time of admission, rather than cognitive impairment.

To date, it remains unclear if cognitive function may influence upper limb functional outcome in patients with stroke who undergo VR intervention and conventional therapy. VR intervention requires the capability of information identification and task execution during the training process [11,25]. This study aimed at exploring the impact of cognitive function on upper limb functional outcome after VR intervention in patients with subacute stroke.

Methods

Study Design

This was a prospective, single-blind, controlled trial, including 2 groups that were divided into 4 subgroups. Participants were first stratified into cognitively normal (CN) if the Mini-Mental State Examination (MMSE) score was greater than or equal to 27, and cognitively impaired (CI) if it was lower than 27 in accordance with previous studies [22,26,27]. Participants were then randomly allocated to either the VR intervention group or the control group, which gave a total of 4 subgroups: cognitively normal VR intervention group (CN_{VR}), cognitively normal control group (CN_{CG}), cognitively impaired VR intervention group (CI_{VR}), and cognitively impaired control group (CI_{CG}). Allocation sequence was randomly generated by a computer program.

Participants and Recruitment

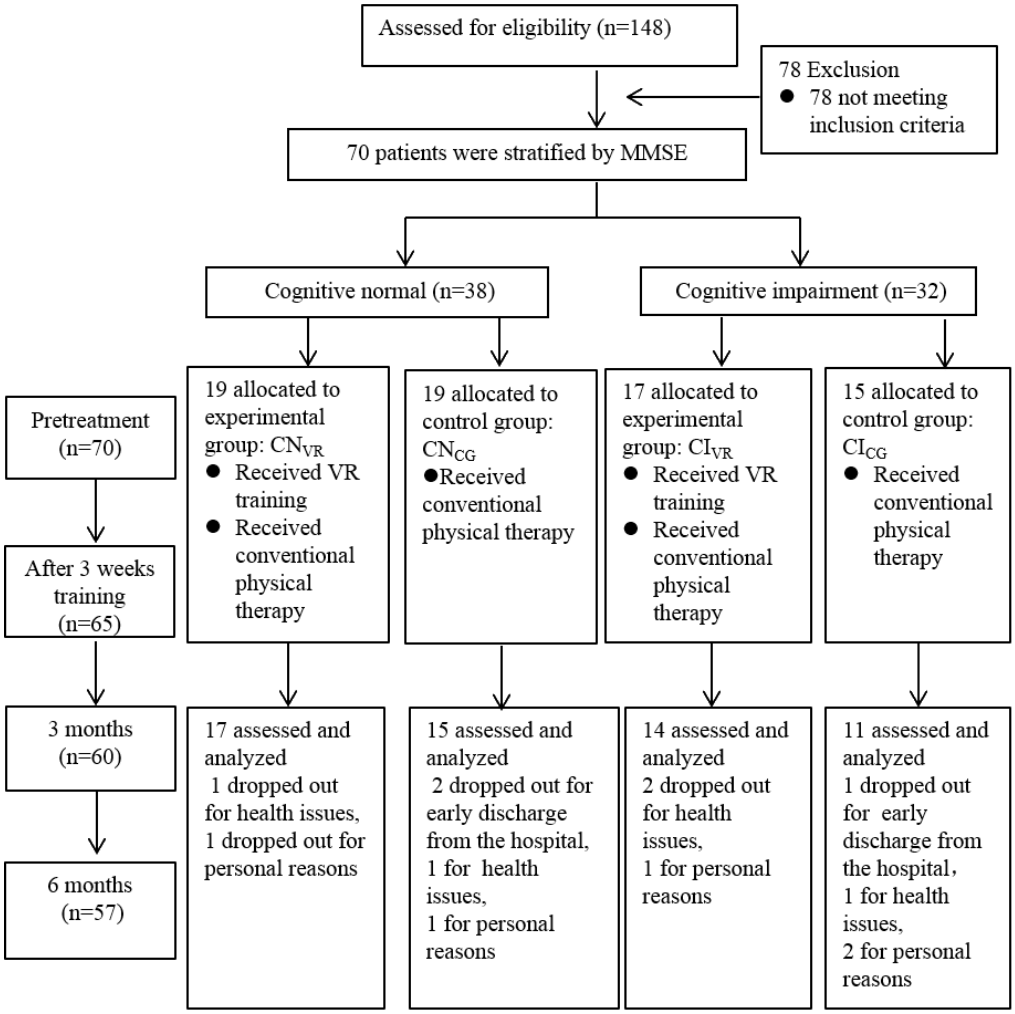
Participants were recruited from the inpatient ward between August 2008 and December 2017. Patients were screened for eligibility as part of routing assessment. Suitable participants were identified by the clinical team and given written information about the study. Participants who were interested to take part were asked to approach a member of the research team. Patients were included if they (1) had the first ever occurrence of unilateral cerebral infarction as confirmed by magnetic resonance imaging or computed tomography; (2) their initial onset was less than 6 months; (3) were able to sit independently for at least 30 minutes; (4) were able to complete the Chinese version of the MMSE assessment individually; (5) their education level is high-to-middle school (including technical secondary school) or above; (6) had no auditory and visual disorder; (7) were dextromanual; (8) were able to provide written informed consent. Patients were excluded if they had (1) uncontrolled medical conditions such as blood pressure or angina; (2) musculoskeletal impairments or fracture of the extremity; (3) a history of psychological conditions; or (4) were unwilling to participate in the study. Baseline information on anthropometric characteristics, including age, gender, hemiplegic side, and the Brunnstrom stage of upper extremity was collected prior to the experiment as part of routine clinical assessment.

A total of 148 patients were screened for suitability, of which 70 patients met the inclusion criteria. As many as 38 and 32

XSL•FO
RenderX

participants were classified as cognitively normal and cognitively impaired, respectively; 13 participants did not complete the study due to personal reasons, health insurance issues, and unexpected early discharge from the hospital. Besides, 2 participants reported slight fatigue after VR training, and 1 complained of pain in shoulder, which resulted in the stopping of VR training immediately. Figure 1 illustrates the number of participants at each stage of the study.

Figure 1. The number of participants at each stage of the study.



Intervention

Overview

The VR intervention group consisted of a 30-minute session of conventional rehabilitation program, followed by a 30-minute nonimmersive VR training using Microsoft Xbox 360 Kinect (Microsoft Corporation). Participants in the control group received 1-hour of conventional rehabilitation program. Both the VR group and the control group received treatment 5 days a week for a total of 3 weeks.

VR Intervention

This study used the Microsoft Xbox 360 Kinect to provide non-VR intervention. The motion camera monitors the body and extremity movements in 3D and tracks movements in real time. Participants were positioned 1.5-3 m away from the Kinect sensor. The following games were used for VR training: Balloon Buster, Table Tennis, Bowling, and Traffic Control. For participants who were at Brunnstrom stage 3 or above, bilateral shoulder and elbow movements were performed actively in the direction of abduction, adduction, flexion, and extension. For participants who were below stage 3 of the Brunnstrom classification, the unaffected arm may assist the activity of the affected arm. If the participant reported fatigue, abnormalities in breathing, or complained of pain, training was stopped immediately. Participants were informed about the experimental procedure. The operating procedure of the training device was demonstrated by the physiotherapist prior to start of the training session.

Conventional Therapy

The conventional physical therapy regimen for upper limb function included range of motion exercises, muscle strengthening, functional training, neurodevelopmental treatment, proprioceptive neuromuscular facilitation, and electrotherapy. The specific rehabilitation tasks for each participants were determined based on clinical needs as deemed appropriate by the treating clinicians.

XSL•FO
RenderX

Outcome Measures

Outcome measures were recorded on 4 occasions: before the intervention, 3 weeks after the intervention, 3 and 6 months after the intervention. A licensed physiotherapist who was blinded to participants' allocation and was not involved in the intervention program recorded the outcome measures at all measuring time points. Upper limb motor function was assessed by the Fugl-Meyer Assessment for Upper Extremity (FMA-UE) [28]. Barthel Index (BI) and Instrumental Activities of Daily Living (IADL) scales were adopted to assess the level of independence in ADLs [4,29]. All interventions were provided by clinical rehabilitation staff who were not blinded to group allocation.

Statistical Analysis

Descriptive statistics were calculated to describe the data set of all variables. Independent *t* tests and chi-square tests were conducted to compare the demographic data and clinical characteristics between the VR group and the control group at baseline. Repeated measures analysis of variance (ANOVA) and the corresponding nonparametric test were conducted to compare the demographic data and clinical characteristics between the 4 subgroups at baseline. Repeated-measures ANOVA was used to compare the outcomes of FMA-UE, BI, and IADL scores of the 4 subgroups at different time points. The ANOVA results were adjusted using Bonferroni post hoc test if the interaction effects reach significant level (*P*=.05). The level of significance was set at an α level of .05. All statistical analyses were performed using SPSS version 22 (IBM, Inc.).

Ethics Approval

The study was conducted at the Rehabilitation Department of a local hospital, and approved by the Medical Ethical Committee of the First Affiliated Hospital of Sun Yat-sen University (approval no.: 201488). All procedures were conducted in accordance with the Declaration of Helsinki. Participants were informed that they had an equal chance to be allocated to either the VR intervention or the conventional therapy group. Written consent was obtained from all participant prior to study enrollment. No changes were made to the planned methods after trial commencement. All mandatory laboratory health and safety procedures were complied with during the course of the study.

Availability of Data and Materials

The data set supporting the conclusions of this article is available from the authors upon request.

Results

Demographic characteristics of the patients in the VR group and the control group are presented in Table 1. No unintended effect was reported by participants of both groups. No statistically significant difference was also found between the VR group and the control group in terms of age (*P*=.85), stroke onset time (*P*=.10), MMSE score (*P*=.81), National Institute of Health Stroke Scale (NIHSS; *P*=.55) score, gender (*P*=.32), hemiplegic side (*P*=.44), and Brunnstrom stages of upper extremity (arm: *P*=.63; hand: *P*=.73). Table 2 presents the clinical characteristics of each subgroup. For NIHSS, the ANOVA showed significant interaction effects (treatment × cognitive effect *P*=.006). Bonferroni post hoc test indicated a significant difference between the CNVR subgroup and the CIVR subgroup (*P*=.004).

Table 1. Demographic characteristics of the VR group and the control group.

Variables	Experimental VR[a] group (n=31)	Control group (n=26)	*P* value
Age (years), mean (SD)	59.25 (10.70)	59.12 (11.62)	.85
Disease duration (days), mean (SD)	43.42 (40.41)	30.15 (15.07)	.10
MMSE[b], mean (SD)	25.32 (4.70)	25.00 (5.24)	.81
NIHSS[c], mean (SD)	7.10 (3.29)	6.58 (3.25)	.55
Sex (female/male), n	6/25	8/18	.32
Hemiplegic side (right/left), n	13/18	14/12	.44
Brunnstrom stage of arm (stage 1-6), n	5/11/8/3/3/1	2/13/3/3/4/1	.63
Brunnstrom stage of hand (stage 1-6), n	5/16/2/4/2/2	5/11/2/2/5/1	.73

[a]VR: virtual reality.

[b]MMSE: Mini-Mental State Examination.

[c]NIHSS: National Institute of Health Stroke Scale.

XSL•FO
RenderX

Table 2. Demographic characteristics of 4 subgroups of the stroke survivors.

Variables	CN_{VR}[a] subgroup (n=17)	CI_{VR}[b] subgroup (n=14)	CN_{CG}[c] subgroup (n=15)	CI_{CG}[d] subgroup (n=11)	ANOVA[e] (treatment)		ANOVA (cognitive)		ANOVA (treatment × cognitive)		Chi-square test	
					$F_{1,53}$	*P* value	$F_{1,53}$	*P* value	$F_{1,53}$	*P* value	χ^2_3	*P* value
Age (years), mean (SD)	57.29 (11.08)	62.57 (9.57)	54.40 (10.37)	65.55 (10.40)	<0.001	.99	8.653	.005	1.105	.30		
Disease duration (days), mean (SD)	36.47 (36.99)	51.86 (43.88)	28.80 (14.32)	32.00 (16.57)	2.656	.11	1.211	.28	0.521	.47		
MMSE[f], mean (SD)	28.71 (1.05)	21.21 (4.06)	28.60 (1.12)	20.09 (4.57)	0.607	.44	102.844	<.001	0.416	.52		
NIHSS[g], mean (SD)	5.35 (2.55)	9.21 (2.86)	6.87 (3.25)	6.18 (3.37)	0.901	.35	3.940	.052	8.070	.006[h,i]		
Sex (female/male), n	2/15	4/10	5/10	3/8							2.291	.54
Hemiplegic side (right/left), n	7/10	6/8	6/9	8/3							3.540	.32
Brunnstrom stage of arm (stage 1-6), n	1/5/4/3/3/1	4/6/4/0/0/0	1/10/1/2/1/0	1/3/2/1/3/1							18.683	.23
Brunnstrom stage of hand (stage 1-6), n	0/9/0/4/3/1	5/7/2/0/0/0	2/9/2/0/2/0	3/2/0/2/3/1							25.014	.050

[a]CN_{VR}: cognitive normal virtual reality group.

[b]CI_{VR}: cognitive impaired virtual reality group.

[c]CN_{CG}: cognitive normal control group.

[d]CI_{CG}: cognitive impaired control group.

[e]ANOVA: analysis of variance.

[f]MMSE: Mini-Mental State Examination.

[g]NIHSS: National Institute of Health Stroke Scale.

[h]ANOVA result was significant.

[i]Bonferroni post hoc test was significant between the CNVR subgroup and the CIVR subgroup (*P*=.004).

The group comparisons between the VR group and the control group indicated significant improvements in FMA-UE (treatment effect *P*=.67; time effect *P*<.001), BI (treatment effect *P*=.39; time effect *P*<.001), and IADL (treatment effect *P*=.97, time effect *P*<.001) in both the VR and control groups after the intervention (Table 3). Bonferroni post hoc test was not conducted to adjust the statistical result as no interaction effects were observed in ANOVA.

For the comparisons between the 4 subgroups, repeated measures ANOVA indicated that FMA-UE scores (time effect *P*<.001), BI scores (time effect *P*<.001), and IADL scores (time effect *P*<.001) were significantly different in each subgroup after the intervention. Table 4 presents the results of repeated measures ANOVA of the 4 subgroups. For FMA-UE and IADL scores, Bonferroni post hoc test was not conducted as the ANOVA results showed no interaction effects. For the BI score, the ANOVA results showed significant interaction effects (treatment × time × cognitive effect *P*=.04). Bonferroni post hoc tests indicated statistically significant differences between the CN_{VR} subgroup and the CI_{VR} subgroup at each measuring time point (Pretreatment *P*=.002; 3 weeks *P*=.005; 3 months *P*=.01; 6 months *P*=.03). There was also a statistically significant difference between the CI_{CG} subgroup and the CI_{VR} subgroup at baseline (*P*<.001).

XSL•FO
RenderX

Table 3. Clinical parameters of the experimental VR group and the control group before and after the intervention.

Parameter	Experimental VR[a] group (n=31)	Control group (n=26)
FMA-UE[b] score[c], mean (SD); (range)		
Pretreatment	27.68 (18.29); 4-63	27.69 (19.92); 6-65
3 weeks' treatment	31.87 (19.42); 8-64	31.54 (19.96); 10-66
3 months	35.90 (21.19); 8-66	35.88 (19.95); 12-66
6 months	38.06 (21.48); 13-66	38.19 (19.01); 14-66
BI[d] score[e], mean (SD); (range)		
Pretreatment	50.65 (25.10); 10-100	60.77 (24.85); 20-100
3 weeks' treatment	64.52 (26.86); 25-100	69.62 (24.41); 25-100
3 months	72.90 (22.67); 30-100	75.96 (23.20); 25-100
6 months	77.90 (20.55); 30-100	81.73 (19.95); 35-100
IADL[f] score[g], mean (SD); (range)		
Pretreatment	2.65 (1.99); 0-8	2.46 (2.13); 1-7
3 weeks' treatment	3.16 (2.30); (0-8)	3.12 (1.97); 1-7
3 months	3.90 (2.53); 0-8	3.69 (2.33); 1-8
6 months	4.52 (2.94); 0-8	4.73 (2.28); 1-8

[a]VR: virtual reality.

[b]FMA-UE: Fugl-Meyer Assessment for Upper Extremity.

[c]Two-way ANOVA: time, $P<.001$ (significant); treatment, $P=.99$; treatment × time, $P=.95$.

[d]BI: Barthel Index.

[e]Two-way ANOVA: time, $P<.001$ (significant); treatment, $P=.34$; treatment × time, $P=.32$.

[f]IADL: instrumental activities of daily living.

[g]Two-way ANOVA: time, $P<.001$ (significant); treatment, $P=.92$; treatment × time, $P=.81$.

XSL•FO
RenderX

Table 4. Clinical parameters of the CNVR, CIVR, CNVR, and CNCG subgroups before and after the intervention.

	CN_{VR}[a] subgroup (n=17)	CI_{VR}[b] subgroup (n=14)	CN_{CG}[c] subgroup (n=15)	CI_{CG}[d] subgroup (n=11)
FMA-UE[e] scores, mean (SD); range[f]				
Pretreatment	33.23 (18.76); 4-63	20.93 (15.79); 6-58	23.20 (17.65); 6-61	33.82 (22.03); 9-65
3 weeks' treatment	39.47 (18.91); 8-64	22.64 (16.22); 8-58	26.07 (16.65); 11-60	39.00 (22.39); 10-66
3 months	43.58 (18.65); 14-66	22.64 (16.22); 8-58	31.47 (17.43); 12-66	41.91 (22.37); 12-66
6 months	46.35 (18.94); 14-66	28.00 (20.58); 5-65	34.20 (17.08); 14-66	43.63 (20.93); 14-66
BI[g] scores, mean (SD); range[h]				
Pretreatment	63.24 (25.98); 25-100	35.35 (14.61); 10-60	57.33 (24.04); 20-100	65.45 (26.31); 25-100
3 weeks' treatment	76.18 (23.15); 30-100	50.36 (25.83); 25-100	70.33 (23.94); 25-100	68.64 (26.18); 25-100
3 months	82.35 (17.42); 45-100	61.43 (24.37); 30-100	77.67 (20.78); 35-100	73.64 (27.02); 25-100
6 months	85.00 (17.68); 55-100	69.29 (21.83); 30-100	83.33 (20.59); 35-100	79.55 (19.81); 35-100
IADL[i] scores, mean (SD); range[j]				
Pretreatment	3.82 (1.85); 2-8	1.21 (0.97); 0-3	2.80 (1.90); 1-7	2.00 (2.53); 0-7
3 weeks' treatment	4.35 (1.90); 2-8	1.71 (1.90); 0-5	3.47 (1.88); 1-7	2.64 (2.16); 1-7
3 months	4.94 (1.87); 2-8	2.64 (2.71); 0-8	3.67 (2.09); 1-7	3.73 (2.83); 1-8
6 months	5.71 (2.23); 2-8	3.07 (3.12); 0-8	5.27 (1.94); 2-8	4.00 (2.68); 1-8

[a]CN_{VR}: cognitive normal virtual reality group.

[b]CI_{VR}: cognitive impaired virtual reality group.

[c]CN_{CG}: cognitive normal control group.

[d]CI_{CG}: cognitive impaired control group.

[e]FMA-UE: Fugl-Meyer assessment for Upper Extremity.

[f]Repeated measures analysis of variance: Time, $F_{3,159}$=52.398, P<.001 (significant); treatment, $F_{1,53}$=0.099, P=.76; cognitive, $F_{1,53}$=0.281, P=.60. Time × treatment, $F_{3,159}$=0.013, P=.97; time × cognitive, $F_{3,159}$=1.576, P=.22; treatment × cognitive, $F_{1,53}$=7.358, P=.009 (significant). Treatment × time × cognitive, $F_{3,159}$=1.329, P=.27.

[g]BI: Barthel Index.

[h]Repeated measures analysis of variance: Time, $F_{3,159}$=49.619, P<.001 (significant); treatment, $F_{1,53}$=1.446, P=.24; cognitive, $F_{1,53}$=4.372, P=.004 (significant). Time × treatment, $F_{3,159}$=1.687, P=.19; time × cognitive, $F_{3,159}$=0.453, P=.66; treatment × cognitive, $F_{1,53}$=4.111, P=.05 (significant). Treatment × time × cognitive, $F_{3,159}$=3.161, P=.04 (significant). The Bonferroni post hoc test was significant between the CNVR subgroup and the CIVR subgroup (P<.001).

[i]IADL: Instrumental Activities of Daily Living.

[j]Repeated measures ANOVA: Time, $F_{3,159}$=20.051, P<.001 (significant); treatment, $F_{1,53}$=0.001, P=.98; cognitive, $F_{1,53}$=11.807, P=.001 (significant). Time × treatment, $F_{3,159}$=0.176, P=.87; time × cognitive, $F_{3,159}$=0.808, P=.47; treatment × cognitive, $F_{1,53}$=3.758, P=.06. Treatment × time × cognitive, $F_{3,159}$=0.262, P=.80.

Discussion

Principal Findings

This study aimed to explore if cognitive status may influence upper limb functional outcomes in patients with stroke who underwent VR intervention and conventional therapy. This study observed no significant difference in all outcome measures between the VR intervention group and the conventional therapy group. The improvements in upper limb function and independency in performing activities of daily living persisted through to 3 and 6 months after the intervention in all subgroups. The improvement of BI scores in those with impaired cognitive function was higher than in those who were cognitively normal immediately after the VR intervention.

Functional Activities

The results of this study indicated that participants with impaired cognitive function may gain more improvement in BI score upon undergoing VR intervention, especially 3 weeks after the intervention. VR intervention is a task-oriented tool with cognitive participation training. It requires participants to understand the task to be performed and respond accordingly in the virtual environment. It places more demand on movement anticipation and execution than conventional therapy [30]. The interactions between cognitive, motor function, and VR training are complex and are only partially understood in patients with stroke [7]. Oneş et al [31] reported a significant positive association between cognitive condition at the time of admission and functional outcomes at discharge, indicating the potential

XSL•FO
RenderX

role of cognitive function in motor function recovery. Heruti et al [22] also reported that in addition to the positive correlation between cognitive status (assessed by the MMSE) and Functional Independence Measures (FIMs), participants with better cognitive function had shorter length of stay during hospitalization. When the effect of the VR intervention and conventional therapy in participants with impaired cognitive function was compared, the largest increase in BI score was observed in the CI_{VR} group from baseline to 3 weeks after the intervention. This may suggest that VR intervention may contribute to functional improvement at a faster rate than conventional therapy in patients with stroke with impaired cognitive function. A previous study that investigated the effect of VR intervention on cognitive function and lower limb function also reported greater improvements in cognitive abilities along with significant improvement in activities of daily living of the Functional Independence subscales, despite similar improvements in limb function observed between the VR group and the non-VR group [32]. These findings indicated that patients with stroke with cognitive impairment may gain more benefit from the VR intervention to improve functional activities. However, these benefits appeared to be limited to activities that were more related to basic self-care, rather than to the more complex functional activities that involved the interaction with the outside environment. This was supported by the results of our study that the largest improvement was observed in BI scores, rather than in IADL scores, in the 3 weeks after the intervention in the CI_{VR} group. BI and IADL are 2 fundamentally different aspects of functions. BI refers to functional ability to perform basic self-care activities such as toilet use, grooming, feeding, and walking, whereas the IADL scale assesses aspects such as transport, traveling, and social activities. Thus, any underlying impairments in motor and cognitive function may affect the performance of the IADL task to a greater degree than basic self-care activities [33].

Upper Limb Motor Function

The mechanism of VR intervention was proposed to induce reorganization of the cerebral cortex [30,34,35]. An fMRI study published by our research group reported cortical reorganization of the contralateral sensorimotor cortex [6] after an intervention using Microsoft Xbox 360 Kinect. Positive outcomes for improving upper extremity motor function after the Microsoft Xbox 360 Kinect intervention were subsequently reported [9,29,36]. These findings are consistent with the data of our study, which also reported positive outcomes of upper limb motor function in the VR group. The group comparison between VR intervention and conventional therapy did not reveal significant differences in upper limb motor function and functional improvements. This finding is consistent with a published Cochrane review that concluded no significant difference between VR intervention and conventional therapy in upper limb functional outcome [3]. This finding, however, contradicts with some studies that reported VR intervention is superior to conventional therapy [3,37] in promoting upper limb functional recovery. A possible reason for the observed lack of significant difference between VR and conventional intervention was the matched intervention time in both groups. A Cochrane review reported that studies that reported superior outcome of VR tended to adopt VR as an augmentation to usual dosage of therapy where participants received more treatment time than the control group [3]. Thus, it could not be ruled out that the higher improvement may be related to the higher training dosage.

A previously published study proposed that cognitive status on admission correlated with motor outcome [22]. This study, however, did not observe any significant difference in FMA score after the intervention between participants who were cognitively normal or cognitively impaired at any of the recording time points. The findings of this study suggested that VR intervention is as effective as conventional therapy in improving upper limb motor function for patients with stroke who were cognitively impaired. These contradicting findings may be related to the diffused assessment of cognitive function of the MMSE [26] and the difference in the instrument adopted to assess motor function. MMSE broadly incorporates several categories such as time and place orientation, short-term memory, attention, recall, calculation, language, and visual spatial abilities [26], which may not be directly related to upper limb motor function. More recently published studies proposed that it was specifically the executive dysfunctions of cognitive impairment that may be related to motor function [38]. The study by Heruti et al [22] adopted the motor function component of the FIM and noted that the absolute gain of motor function may not be related to cognitive status, but is more related to the outcome of functional activities, which the FIM motor score was based on. This suggestion is given further support from a recent study that compared the effect of neurocognitive robot-assisted rehabilitation [39]. The study reported no significant difference in upper limb motor function (assessed by FMA-UE) and cognitive function (assessed by MMSE) gain was observed between neurocognitive robot-assisted rehabilitation and conventional therapy. This finding provide further evidence to suggest that cognitive status, as assessed by the MMSE, may not play an influencing role in the outcome of upper limb motor function provided by VR intervention [40].

Limitations

The findings of this study should be interpreted with caution due to its limitations. There are insufficient data regarding the duration and frequency of treatment administered by the Microsoft Xbox 360 Kinect system. The training games were not specifically designed for stroke rehabilitation, which may contribute to the underestimation of the benefit of the VR intervention. The restricted hospitalization period (3 weeks) placed a limitation on treatment intensity. The sample size of each subgroup was small due to attrition during the follow-up period. The data therefore were likely to contain type II errors. Ikbali Afsar et al [11] proposed that the sensitivity of the Fugl-Meyer Scale could be affected by inadequate sample size, short duration of treatment, and short follow-up period. Future randomized trials with a large sample of subgroup population should be conducted to verify the findings of this study.

Conclusion

In conclusion, VR-based intervention training is as effective as traditional conventional therapy regardless of cognitive status. Patients with stroke who have impaired cognitive function may

XSL•FO
RenderX

gain more improvement in upper limb function that is related to self-care activity immediately after the VR intervention. Findings of this study adds further support that cognitive function plays an important role in upper limb motor function recovery. Clinicians who consider offering VR intervention for this pathological group should take into consideration the cognitive status to gain optimal benefit of the rehabilitation program. The presence of cognitive impairment should be considered when planning individualized rehabilitation program for patients with stroke to maximize upper limb function recovery.

Acknowledgments

The authors thank all the study participants. Special thanks must be given to Dr Tingni Li for the statistical support and the profound scientific discussion on the research findings. This study is supported by the National Key Research and Development Program of China (grant 2020YFC2004304), Natural Science Foundation of China (grants 31771016, 32071316, and 81971224), Guangdong Basic and Applied Basic Research Foundation (grant 2020A1515011356), 5010 Planning Project of Sun Yat-Sen University of China (grant 2014001), Medical Scientific Research Foundation of Guangdong Province of China (grant A2021334), the Fundamental Research Funds for the Central Universities (grant G2022WD01006， W016204), and the Key Research and Development Project of Shaanxi province (2022SF-117).

Authors' Contributions

All authors have read and approved the final manuscript. All authors meet the 4 primary ICMJE criteria for authorship. In addition, all authors have been actively involved in the study in different capacities: YL and WLAL were involved in all stages of the study including data collection, analysis, interpretation, and drafting of the manuscript. YRM, RB, JLZ, and ZX were responsible for the design of the research protocol, data collection, and analysis. LL and DFH revised the manuscript, interpreted the data, and managed the trial.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 436 KB - games_v10i3e33755_app1.pdf]

References

1. Murray C, Vos T, Lozano R, Naghavi M, Flaxman A, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4]
2. Brauer SG, Hayward KS, Carson RG, Cresswell AG, Barker RN. The efficacy of SMART Arm training early after stroke for stroke survivors with severe upper limb disability: a protocol for a randomised controlled trial. BMC Neurol 2013 Jul 02;13:71 [FREE Full text] [doi: 10.1186/1471-2377-13-71] [Medline: 23815739]
3. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
4. Saposnik G, Cohen LG, Mamdani M, Pooyania S, Ploughman M, Cheung D, Stroke Outcomes Research Canada. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. Lancet Neurol 2016 Sep;15(10):1019-1027 [FREE Full text] [doi: 10.1016/S1474-4422(16)30121-1] [Medline: 27365261]
5. Ahn S, Hwang S. Virtual rehabilitation of upper extremity function and independence for stoke: a meta-analysis. J Exerc Rehabil 2019 Jun;15(3):358-369 [FREE Full text] [doi: 10.12965/jer.1938174.087] [Medline: 31316927]
6. Bao X, Mao Y, Lin Q, Qiu Y, Chen S, Li L, et al. Mechanism of Kinect-based virtual reality training for motor functional recovery of upper limbs after subacute stroke. Neural Regen Res 2013 Nov 05;8(31):2904-2913 [FREE Full text] [doi: 10.3969/j.issn.1673-5374.2013.31.003] [Medline: 25206611]
7. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. J Neuroeng Rehabil 2018 Mar 27;15(1):29 [FREE Full text] [doi: 10.1186/s12984-018-0370-2] [Medline: 29587853]
8. Perez-Marcos D, Chevalley O, Schmidlin T, Garipelli G, Serino A, Vuadens P, et al. Increasing upper limb training intensity in chronic stroke using embodied virtual reality: a pilot study. J Neuroeng Rehabil 2017 Nov 17;14(1):119 [FREE Full text] [doi: 10.1186/s12984-017-0328-9] [Medline: 29149855]

XSL•FO
RenderX

9. Sin H, Lee G. Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia. Am J Phys Med Rehabil 2013 Oct;92(10):871-880. [doi: 10.1097/PHM.0b013e3182a38e40] [Medline: 24051993]
10. Yin CW, Sien NY, Ying LA, Chung SFCM, Tan May Leng D. Virtual reality for upper extremity rehabilitation in early stroke: a pilot randomized controlled trial. Clin Rehabil 2014 Nov;28(11):1107-1114. [doi: 10.1177/0269215514532851] [Medline: 24803644]
11. Ikbali Afsar S, Mirzayev I, Umit Yemisci O, Cosar Saracgil SN. Virtual Reality in Upper Extremity Rehabilitation of Stroke Patients: A Randomized Controlled Trial. J Stroke Cerebrovasc Dis 2018 Dec;27(12):3473-3478. [doi: 10.1016/j.jstrokecerebrovasdis.2018.08.007] [Medline: 30193810]
12. Lee G. Effects of training using video games on the muscle strength, muscle tone, and activities of daily living of chronic stroke patients. J Phys Ther Sci 2013 May;25(5):595-597 [FREE Full text] [doi: 10.1589/jpts.25.595] [Medline: 24259810]
13. Varghese JP, Merino DM, Beyer KB, McIlroy WE. Cortical control of anticipatory postural adjustments prior to stepping. Neuroscience 2016 Jan 28;313:99-109. [doi: 10.1016/j.neuroscience.2015.11.032] [Medline: 26608123]
14. Denti L, Agosti M, Franceschini M. Outcome predictors of rehabilitation for first stroke in the elderly. Eur J Phys Rehabil Med 2008 Mar;44(1):3-11 [FREE Full text] [Medline: 18385622]
15. Schaefer SY, Haaland KY, Sainburg RL. Hemispheric specialization and functional impact of ipsilesional deficits in movement coordination and accuracy. Neuropsychologia 2009 Nov;47(13):2953-2966 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2009.06.025] [Medline: 19573544]
16. Lindberg PG, Gäverth J, Islam M, Fagergren A, Borg J, Forssberg H. Validation of a new biomechanical model to measure muscle tone in spastic muscles. Neurorehabil Neural Repair 2011 Sep;25(7):617-625. [doi: 10.1177/1545968311403494] [Medline: 21490269]
17. Kalaska JF. From intention to action: motor cortex and the control of reaching movements. Adv Exp Med Biol 2009;629:139-178. [doi: 10.1007/978-0-387-77064-2_8] [Medline: 19227499]
18. Chen L, Mao Y, Ding M, Li L, Leng Y, Zhao J, et al. Assessing the Relationship Between Motor Anticipation and Cortical Excitability in Subacute Stroke Patients With Movement-Related Potentials. Front Neurol 2018;9:881 [FREE Full text] [doi: 10.3389/fneur.2018.00881] [Medline: 30386292]
19. Dean PJA, Seiss E, Sterr A. Motor planning in chronic upper-limb hemiparesis: evidence from movement-related potentials. PLoS One 2012;7(10):e44558 [FREE Full text] [doi: 10.1371/journal.pone.0044558] [Medline: 23049676]
20. Hao Z, Wang D, Zeng Y, Liu M. Repetitive transcranial magnetic stimulation for improving function after stroke. Cochrane Database Syst Rev 2013 May 31(5):CD008862 [FREE Full text] [doi: 10.1002/14651858.CD008862.pub2] [Medline: 23728683]
21. Barker-Collo S, Feigin V. The impact of neuropsychological deficits on functional stroke outcomes. Neuropsychol Rev 2006 Jun;16(2):53-64. [doi: 10.1007/s11065-006-9007-5] [Medline: 16967344]
22. Heruti RJ, Lusky A, Dankner R, Ring H, Dolgopiat M, Barell V, et al. Rehabilitation outcome of elderly patients after a first stroke: effect of cognitive status at admission on the functional outcome. Arch Phys Med Rehabil 2002 Jun;83(6):742-749. [doi: 10.1053/apmr.2002.32739] [Medline: 12048650]
23. Paker N, Buğdaycı D, Tekdöş D, Kaya B, Dere C. Impact of cognitive impairment on functional outcome in stroke. Stroke Res Treat 2010 Mar 11;2010:652612 [FREE Full text] [doi: 10.4061/2010/652612] [Medline: 20798755]
24. Diamond PT, Felsenthal G, Macciocchi SN, Butler DH, Lally-Cassady D. Effect of cognitive impairment on rehabilitation outcome. Am J Phys Med Rehabil 1996;75(1):40-43. [doi: 10.1097/00002060-199601000-00011] [Medline: 8645438]
25. Mumford N, Wilson PH. Virtual reality in acquired brain injury upper limb rehabilitation: evidence-based evaluation of clinical research. Brain Inj 2009 Mar;23(3):179-191. [doi: 10.1080/02699050802695566] [Medline: 19205954]
26. Folstein M, Folstein S, McHugh P. ?Mini-mental state?: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 1975:A [FREE Full text] [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
27. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res 2009 Jan;43(4):411-431. [doi: 10.1016/j.jpsychires.2008.04.014] [Medline: 18579155]
28. Page SJ, Levine P, Hade E. Psychometric properties and administration of the wrist/hand subscales of the Fugl-Meyer Assessment in minimally impaired upper extremity hemiparesis in stroke. Arch Phys Med Rehabil 2012 Dec;93(12):2373-6.e5 [FREE Full text] [doi: 10.1016/j.apmr.2012.06.017] [Medline: 22759831]
29. Turolla A, Dam M, Ventura L, Tonin P, Agostini M, Zucconi C, et al. Virtual reality for the rehabilitation of the upper limb motor function after stroke: a prospective controlled trial. J Neuroeng Rehabil 2013 Aug 01;10:85 [FREE Full text] [doi: 10.1186/1743-0003-10-85] [Medline: 23914733]
30. Manuli A, Maggio MG, Latella D, Cannavò A, Balletta T, De Luca R, et al. Can robotic gait rehabilitation plus Virtual Reality affect cognitive and behavioural outcomes in patients with chronic stroke? A randomized controlled trial involving three different protocols. J Stroke Cerebrovasc Dis 2020 Aug;29(8):104994. [doi: 10.1016/j.jstrokecerebrovasdis.2020.104994] [Medline: 32689601]
31. Carod-Artal FJ, González-Gutiérrez JL, Herrero JAE, Horan T, De Seijas EV. Functional recovery and instrumental activities of daily living: follow-up 1-year after treatment in a stroke unit. Brain Inj 2002 Mar;16(3):207-216. [doi: 10.1080/02699050110103337] [Medline: 11874614]

XSL•FO
RenderX

32. Carey JR, Bhatt E, Nagpal A. Neuroplasticity promoted by task complexity. Exerc Sport Sci Rev 2005 Jan;33(1):24-31. [Medline: 15640717]
33. Takeuchi N, Izumi SI. Maladaptive plasticity for motor recovery after stroke: mechanisms and approaches. Neural Plast 2012;2012:359728 [FREE Full text] [doi: 10.1155/2012/359728] [Medline: 22792492]
34. Weiss P. Virtual reality in neurorehabilitation. In: Selzer M, editor. Textbook Neural Repair Rehabilitation (Vol. 51). Cambridge, UK: Cambridge University Press; 2006:198-218.
35. Park DS, Lee DG, Lee K, Lee G. Effects of Virtual Reality Training using Xbox Kinect on Motor Function in Stroke Survivors: A Preliminary Study. J Stroke Cerebrovasc Dis 2017 Oct;26(10):2313-2319. [doi: 10.1016/j.jstrokecerebrovasdis.2017.05.019] [Medline: 28606661]
36. Rogers JM, Duckworth J, Middleton S, Steenbergen B, Wilson PH. Elements virtual rehabilitation improves motor, cognitive, and functional outcomes in adult stroke: evidence from a randomized controlled pilot study. J Neuroeng Rehabil 2019 May 15;16(1):56 [FREE Full text] [doi: 10.1186/s12984-019-0531-y] [Medline: 31092252]
37. Muscari A, Spiller I, Bianchi G, Fabbri E, Forti P, Magalotti D, Pianoro Study Group. Predictors of cognitive impairment assessed by Mini Mental State Examination in community-dwelling older adults: relevance of the step test. Exp Gerontol 2018 Jul 15;108:69-76. [doi: 10.1016/j.exger.2018.03.021] [Medline: 29601853]
38. Laakso HM, Hietanen M, Melkas S, Sibolt G, Curtze S, Virta M, et al. Executive function subdomains are associated with post-stroke functional outcome and permanent institutionalization. Eur J Neurol 2019 Mar;26(3):546-552. [doi: 10.1111/ene.13854] [Medline: 30414288]
39. Ranzani R, Lambercy O, Metzger JC, Califfi A, Regazzi S, Dinacci D, et al. Neurocognitive robot-assisted rehabilitation of hand function: a randomized control trial on motor recovery in subacute stroke. J Neuroeng Rehabil 2020 Aug 24;17(1):115 [FREE Full text] [doi: 10.1186/s12984-020-00746-7] [Medline: 32831097]
40. Molnar FJ, Benjamin S, Hawkins SA, Briscoe M, Ehsan S. One Size Does Not Fit All: Choosing Practical Cognitive Screening Tools for Your Practice. J Am Geriatr Soc 2020 Oct;68(10):2207-2213. [doi: 10.1111/jgs.16713] [Medline: 32720722]

Abbreviations

ANOVA: analysis of variance
BI: Barthel Index
CICG: cognitive impaired control group
CIVR: cognitive impaired virtual reality group
CNCG: cognitive normal control group
CNVR: cognitive normal virtual reality group
FIM: Functional Independence Measure
FMA-UE: Fugl-Meyer assessment for Upper Extremity
IADL: Instrumental Activities of Daily Living
MMSE: Mini-Mental State Examination
NIHSS: National Institute of Health Stroke Scale
TMS: transcranial magnetic stimulation
VR: virtual reality

Edited by N Zary; submitted 22.09.21; peer-reviewed by M Rampioni, PC Wang; comments to author 17.02.22; revised version received 28.03.22; accepted 07.04.22; published 08.07.22.

Please cite as:
Leng Y, Lo WLA, Mao YR, Bian R, Zhao JL, Xu Z, Li L, Huang DF
The Impact of Cognitive Function on Virtual Reality Intervention for Upper Extremity Rehabilitation of Patients With Subacute Stroke: Prospective Randomized Controlled Trial With 6-Month Follow-up
JMIR Serious Games 2022;10(3):e33755
URL: https://games.jmir.org/2022/3/e33755
doi:10.2196/33755
PMID:35802415

©Yan Leng, Wai Leung Ambrose Lo, Yu Rong Mao, Ruihao Bian, Jiang Li Zhao, Zhiqin Xu, Le Li, Dong Feng Huang. Originally published in JMIR Serious Games (https://games.jmir.org), 08.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Interactive Somatosensory Games in Rehabilitation Training for Older Adults With Mild Cognitive Impairment: Usability Study

Chien-Hsiang Chang[1*], MA; Chung-Hsing Yeh[2*], PhD; Chien-Cheng Chang[3*], PhD; Yang-Cheng Lin[1*], PhD

[1]Department of Industrial Design, National Cheng Kung University, Tainan City, Taiwan

[2]Faculty of Information Technology, Monash University, Melbourne, Australia

[3]Department of Industrial Design, National United University, Miaoli, Taiwan

[*]all authors contributed equally

Corresponding Author:
Yang-Cheng Lin, PhD
Department of Industrial Design
National Cheng Kung University
Number 1, University Road
Tainan City, 701
Taiwan
Phone: 886 62757575 ext 54343
Email: lyc0914@mail.ncku.edu.tw

Abstract

Background: In aging societies, dementia risk increases with advancing age, increasing the incidence of dementia-related degenerative diseases and other complications, especially fall risk. Dementia also escalates the care burden, impacting patients, their families, social welfare institutions, and the social structure and medical system.

Objective: In elderly dementia, traditional card recognition rehabilitation (TCRR) does not effectively increase one's autonomy. Therefore, from the usability perspective, we used the Tetris game as a reference to develop an interactive somatosensory game rehabilitation (ISGR) with nostalgic style for elders with mild cognitive impairment (MCI). Through intuitive gesture-controlled interactive games, we evaluated subjective feelings concerning somatosensory game integration into rehabilitation to explore whether the ISGR could improve the willingness to use and motivation for rehabilitation among elders with MCI.

Methods: A total of 15 elders with MCI (7 males and 8 females with an average age of 78.4 years) underwent 2 experiments for 15 minutes. During experiment 1, TCRR was performed, followed by completing the questionnaire of the System Usability Scale (SUS). After 3-5 minutes, the second experiment (the ISGR) was conducted, followed by completing another SUS. We used SUS to explore differences in impacts of TCRR and ISGR on willingness to use among elders with MCI. In addition, we further investigated whether the factor of gender or prior rehabilitation experience would affect the rehabilitation willingness or not.

Results: The novel ISGR made the elderly feel interested and improved their willingness for continuous rehabilitation. According to the overall SUS score, the ISGR had better overall usability performance (73.7) than the TCRR (58.0) (t_{28}=–4.62, P<.001). Furthermore, the ISGR individual item scores of "Willingness to Use" (t_{28}=–8.27, P<.001), "Easy to Use" (t_{28}=–3.17, P<.001), "System Integration" (t_{28}=–5.07, P<.001), and "Easy to Learn" (t_{28}=–2.81, P<.001) were better than TCRR. The somatosensory game was easier to learn and master for females than for males (t_{13}=2.71, P=.02). Besides, the ISGR was easier to use (t_{12}=–2.50, P=.02) and learn (t_{14}=–3.33, P<.001) for those without prior rehabilitation experience. The result indicates that for elders with no rehabilitation experience ISGR was easier to use and simpler to learn than TCRR.

Conclusions: Regardless of prior rehabilitation experience, the ISGR developed in this study was easy to learn and effective in continuously improving willingness to use. Furthermore, the adoption of a nostalgic game design style served the function of cognitive training and escalated interest in rehabilitation. The ISGR also improved user stickiness by introducing different game scenarios and difficulties, increasing long-term interest and motivation for rehabilitation. For future research on the adoption of interactive somatosensory games in rehabilitation, additional rehabilitation movements can be developed to benefit the elderly with MCI.

(JMIR Serious Games 2022;10(3):e38465) doi:10.2196/38465

XSL•FO
RenderX

KEYWORDS

dementia; elderly; usability; gesture recognition; card recognition rehabilitation; interactive somatosensory game

Introduction

Background

Owing to social changes and medical advancement, the global demographic structure is becoming an aging society. The World Health Organization (WHO) [1] has reported that the number of deaths caused by dementia has doubled in recent years, making it the fifth leading cause of death worldwide. At present, there are approximately 50 million people with dementia; besides, the prevalence of dementia among the elderly over 60 years old is 5%-8%. This number is expected to reach 152 million by 2050. Data suggest that the incidence of dementia is increasing annually, which has prompted the WHO to launch the "Global action plan on the public health response to dementia" with reduction in the risk of dementia as one of the priorities. As the size of the local population with dementia rapidly elevates, the Taiwanese government has initiated a new version of the long-term care plan, the "Long-term Care 2.0 in Taiwan," which prioritizes the care of patients with dementia and the establishment of a friendly dementia care model [2]. Song and Wang [3] have indicated that while there are no existing cures for dementia, rehabilitation therapy can be used as a way of daily training. Rehabilitation training, which can be divided into functional, psychological, behavior management, and other activities, such as hand exercises and creative and mind-stimulating activities that promote community communication, is capable of effectively improving the quality of life of patients [4]. Similarly, Fried et al [5] have reported that finger exercise can not only effectively prevent dementia, but also alleviate anxiety as well as enhancing cold hands and feet, stabilizing blood pressure, and improving sleep quality.

Among various treatment methods for dementia, traditional rehabilitation such as card recognition training utilizes commercially available playing cards for cognitive training. However, its long-term use can neither enhance the autonomy of the elderly nor increase user stickiness. On the contrary, digital games have now been introduced as an emerging cognitive training mode and are referred to as "serious games" [6]. Unlike general entertainment games, serious games are called "serious" because they "replace serious themes with interesting games, which are purposeful and helpful to players" [6-9]. With the maturity and popularization of digital technology, increasing attention has been drawn to the application of interactive games in training, believing that it is an effective training tool [10-12]. In terms of game-based training, Zarit et al [13] proposed a combination of training (rehabilitation) and games, which could not only realize the training effect through games, but also improve users' motivation and willingness to use via high interactivity and objectivity.

This study found that existing card-based rehabilitation not only failed to promote the autonomous use among the elderly, but also triggered hand movements only 3-5 times per minute. Therefore, this study developed a nostalgic game (Tetris) combined with gesture detection technology, allowing elders use their finger movements to control the interactive game. Concurrently, the nostalgic game stimulated the elder's memory (returned to their childhood days) as well as increased their frequency of finger movement (8-10 times per minute), thereby effectively improving user motivation as well as achieving a cognitive training effect. However, for the elderly, the operation of interactive somatosensory game rehabilitation (ISGR) was more complicated than that of the traditional card recognition rehabilitation (TCRR). Therefore, whether the operation of ISGR would compromise the willingness to use for elders' rehabilitation remains an unsolved question. Would the factors such as prior rehabilitation experience and gender be affected by their rehabilitation motivation? To answer these questions, the System Usability Scale (SUS) [14] was adopted to investigate the subjective feelings toward the ISGR and willingness to use by elders with mild cognitive impairment [MCI], with the results also compared with the TCRR.

The SUS was developed by Brooke [14] to explore the system-related experience of users, whose feedback can be used as a basis for understanding/improving the usability of the system. Its main purpose is to detect the subjective feelings of users after operating the system and to evaluate and ensure the usability of the system. The SUS collects users' subjective evaluations of products (or systems) by comparing their performances in different tasks and identifies usability and user satisfaction levels, thereby allowing developers to quickly determine the usability of the product (or system) [15]. In addition, according to relevant medical literature, Yen and Bakken [16] proposed the system development life cycle to evaluate the representative expected value and usability for each different development stage [17,18].

This study aimed to develop an interactive somatosensory game to promote the rehabilitation motivation of elders with MCI, to ensure they were satisfied with the ISGR and could continuously use the ISGR. Sections of the paper included discussions of relevant literature on dementia, the design process of the interactive somatosensory game, usability experiments, and analysis and discussion on the SUS.

Literature Review

Overview

In this section, a literature review and analysis of the importance of dementia rehabilitation, the application of gesture recognition to dementia treatment, and the efficacy of nostalgia therapy on cognitive impairment were conducted. Furthermore, this section will discuss the development status of the application of gesture-based interactive somatosensory games in the rehabilitation.

Importance of Dementia Rehabilitation

The cognitive function includes emotion, personality, perception, memory, abstract judgment, and the ability to express things, and MCI refers to the transitional stage during which a normal person develops dementia (increasing with advancing age) [19].

XSL•FO
RenderX

During MCI, people's daily life is basically unimpacted, although their decline in memory exceeds the normal range. While there is currently no standard for the diagnosis of MCI, Petersen et al [20] proposed that individuals should be diagnosed with MCI if they experienced memory impairment and yet had a normal daily life, normal overall cognitive function, normal memory for their age, and no signs of dementia. Studies indicate that about 10%-15% of the patients with MCI could progress to dementia without receiving proper treatment [19].

Despite being a disease, dementia is thought of by many family members as a common aging phenomenon and the importance of its treatment is therefore often neglected. Once the severity of dementia and cognitive impairment elevates, the difficulty of care increases correspondingly. Furthermore, as the patient ages, the risk of fall also rises, with a study reporting that the risk of fall of elderly people with dementia was substantially higher than that of those without [21]. While approximately 20%-30% of the elderly over the age of 65 have experienced falls, that rate is twice as high among the elderly with dementia [22]. Falls not only injure the elderly, but also bring them into a vicious cycle of functional deterioration. Therefore, to prevent dementia from causing additional complications, it is of particular significance to provide the elderly with appropriate cognitive rehabilitation training to improve the cognitive function and delay the deterioration of dementia. Research indicated that cognitive training and rehabilitation methods exerted a positive effect on patients with dementia with cognitive decline. As a nondrug treatment method, rehabilitation not only reduces burdens on the National Health Insurance, but also improves the quality of life and reduces the impact of mental illness on the elderly [23]. Therefore, developing health care for the elderly with dementia as well as delaying and improving the onset of dementia has become a significant social issue that requires immediate attention.

Application of Gesture Recognition in the Treatment of Dementia

The incidence of dementia has been on the rise in recent years, making it a common disease among the elderly. Furthermore, its prevalence increases with advancing age [19]. However, rather than a normal aging phenomenon, dementia is a degenerative disease that is yet to be cured by drug treatment. Despite so, the incidence of dementia can be effectively prevented by activities. Regular exercise has been proven effective in preventing and improving dementia in the elderly [24]. This is because exercise can enhance the function of the cerebral cortex, thereby benefiting brain intelligence. In addition, an appropriate amount of exercise can alleviate mental anxiety, which in turn protects neurons [25,26].

Recent studies have verified the correlation between the fingers and the brain. Fried et al [5] reported a direct relationship between the hands and brain activities. He suggested that while the frontal and temporal cortices in the brain supported human's motor behavior, continuous exercise could affect parietal and frontal regions of the brain that were related to the medial and lateral surface cortices, thereby productively preventing dementia [27]. In addition, Pellegrino and Làdavas [28] indicated that as the control center of various parts of the body, the brain functions through extensive multisensory interactions within a set of interconnected parietal and frontal regions. Taking finger movements as an example, Gardini et al [29] proposed that motor thinking was controlled by a cortical network that mainly involved abstract thinking, cognition, motor control, semantics, and visual image processes.

The book *Fascinating Dermatoglyphics* by Chen [30] mentioned that the left hand corresponds to the right brain, while the right hand corresponds to the left brain. Each finger had different functions and was linked to different areas of the brain. A brief overview of the book [30] is shown in Figure 1.

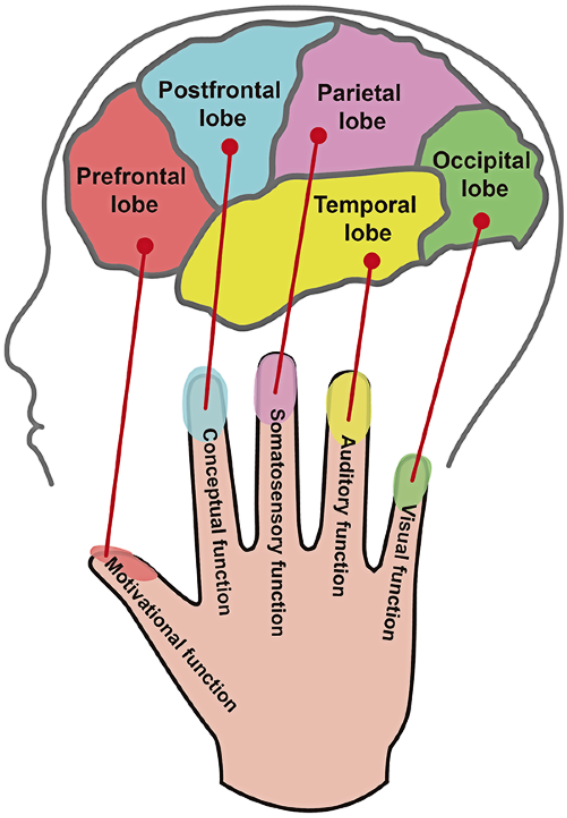
- Thumb: Corresponding to the prefrontal lobe, it affects motivational functions (such as communication management, creative leadership, plan execution, introspective will).
- Index finger: Corresponding to the postfrontal lobe, it affects conceptual functions (such as logical reasoning, computational analysis, language expression, spatial concepts).
- Middle finger: Corresponding to the parietal lobe, it affects somatosensory functions (such as somatosensory recognition, orientation judgment).
- Ring finger: Corresponding to the temporal lobe, it affects auditory functions (such as auditory recognition, emotional control, music rhythm, language understanding).
- Little finger: Corresponding to the occipital lobe, it affects visual functions (such as visual recognition, observation and understanding, image observation).

The book *Stimulating the Thumbs to Make the Brain Younger* by Hasegawa [31] mentioned that moving fingers could increase the blood flow in the motor cortex and somatosensory cortex of the brain by more than 10%, making it effective in preventing dementia [31]. Furthermore, stimulating fingers not only reduces the onset of dementia symptoms, but also improves the blood circulation and motor functions of the human body. In the modern society, our brain constantly works in a stressful environment. Although unaware of it, long-term stress can affect our health. By contrast, finger movements can promote the differentiation and maturation of areas of the brain, thereby exhibiting a positive impact on our health.

With the advancement of science and technology, the cost of production is becoming increasingly lower, making technology products more accessible and their implementation in medical care an important development goal. At present, multiple gesture recognition sensors are commercially available, of which controller-based visual recognition technologies, such as leap motion (LM) and Kinect (Microsoft), are the most common tools for hand rehabilitation. The therapeutic value of these technological tools has been well verified in previous studies [32,33]. Smeragliuolo et al [34] utilized LM in clinical trials and confirmed that it was very sensitive in tracking the hand motion and could provide clinically valuable data.

XSL•FO
RenderX

Figure 1. Areas of the hands linked to different brain functions.



Efficacy of Nostalgia Therapy on Cognitive Impairment

Treatment methods of dementia can be divided into drug treatment and nondrug treatment; the latter includes nostalgia therapy, music therapy, serious training, art therapy, among others [35]. The goal is to improve the quality of life of patients and to prevent and delay the incidence of diseases. Although currently no drug treatment can completely prevent or cure dementia-related degenerative diseases, both drug and nondrug methods are capable of delaying the progression of the disease or alleviating symptoms.

Nostalgia therapy uses a set of planned nursing activities to remind the elderly of people, events, and things that have special meanings to them and guide them to tell their experiences, thereby helping them realize that their life is meaningful and valuable and regain old memories [36]. It can improve the patient's self-esteem and self-confidence as well as alleviating depression. As there are currently limited drugs for dementia and a drug cure is yet to be discovered, nondrug treatment has become the mainstream against dementia. Multiple nondrug treatment methods for dementia are available, among which nostalgia therapy has been shown by many studies to demonstrate positive results on the cognitive function and the quality of life of patients [37]. By satisfying the physical, psychological, and social needs of the elderly, nostalgia therapy can help them realize self-integration and improve their mental status as well as quality of life, thereby not only preventing and reducing the incidence of dementia, but also greatly facilitating the elderly to adapt to their daily life. Therefore, nondrug treatments for dementia have received increasing attention in recent years. Research has confirmed that nondrug therapy can reduce the degree of cognitive degradation and the onset of psychobehavioral symptoms [38]. Furthermore, it can be used to communicate with patients with severe dementia. Therefore, nondrug therapy has been commonly adopted in the care of elderly people and is considered an important measure that maintains the physical and mental health as well as the quality of life of the elderly [38].

Autobiographical memory is a very important memory for humans, the loss of which can lead to a disconnection between the past and the present and more severely, the loss of personal identity [39]. Research has shown that nostalgia therapy exhibits a significant impact on the episodic and semantic memory of the autobiographical memory of patients with MCI [40]. In addition, it positively affects the cognition, communication, and emotion of patients with dementia, thereby improving their quality of life, cognitive ability, and social ability [41]. Duru Asiret and Kapucu [42] recruited 62 patients (31 in the experimental group and 31 in the control group) and asked them to participate in a 35-minute therapy course per week for a duration of 12 weeks before getting tested. The results of their experiment confirmed that memory therapy significantly improved cognition and depression in patients with mild or moderate dementia. Similarly, the Geriatrics Center of the National Cheng Kung University Hospital (in Tainan, Taiwan) established a nostalgic space-time walk path in the clinical environment by decorating the path with nostalgic objects. By walking through the path, the dementia symptoms of the elderly were successfully alleviated (as shown in Figure 2).

XSL•FO
RenderX

Figure 2. Arrangements of nostalgic objects in the Geriatrics Center of the National Cheng Kung University Hospital.



Objective

This study aims to explore the integration of interactive somatosensory games into cognitive training and to investigate the feasibility of applying the interactive somatosensory game developed in this study to the rehabilitation of the elderly.

Methods

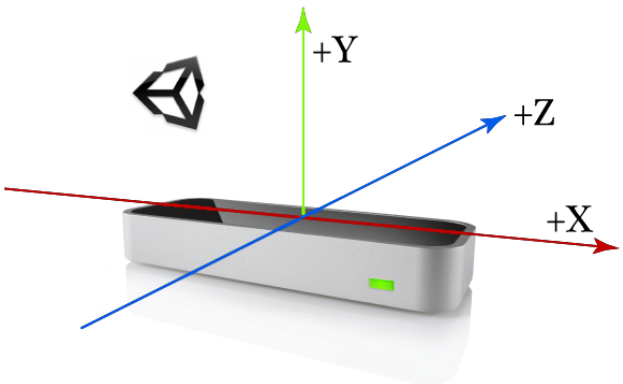
Interactive Somatosensory Game Design

The interactive somatosensory game developed in this study incorporated simple operations and easy-to-learn gameplay, and integrated the gesture recognition technology of LM, which distinguished it from traditional Tetris games. Gesture recognition was used to control the number and the movement of the blocks. This not only made it easier for the elderly to operate, but also facilitated the movement and rehabilitation of their hands through hand exercise, thereby improving the cognitive ability of the elderly and preventing degenerative diseases. The following sections provide an in-depth discussion on the 3 major aspects of the game design, that is, gesture recognition technology, game interface design, and game system design.

Gesture Recognition Technology

With the development of science and technology and the rise of human-computer interaction, motion recognition technology has received substantial attention. As a key technology in motion recognition, gesture recognition captures the dynamic gestures of human hands through visual sensor technology. In recent years, as the development cost has decreased and technology has rapidly developed, many dynamic gesture capturing tools have emerged in the market. This study adopted the LM gesture capture tool, which realizes gesture recognition through dual cameras and 3 infrared LEDs (Figure 3). LM can not only recognize multiple hands simultaneously, but also has a submillimeter accuracy and a sensing distance of 25-600 mm above the device [43], making it the most accurate economic gesture recognition tool in the market [44]. The biggest difference between LM and other hand recognition tools such as Microsoft Kinect is that LM is capable of capturing depth data. By calculating the palm center position and other related point positions through the palm orientation and fingertip position, LM can acquire depth information without additional calculations [45]. LM can be utilized in various gesture recognition applications, such as gesture control interface, acquisition of user's hand skeletons. Therefore, this study adopted LM as the gesture recognition technology. The Unity engine was also utilized to integrate technical tools into game development.

Figure 3. Recognition of hand orientation by the leap motion device.



Game Interface Design

The game interface design developed in this study was presented in a nostalgic style, providing players with a unique experience through cultural scenes of Taiwan in the 1960s, such as traditional Taiwanese grocery stores, classic movie posters. The background of this game was based on nostalgic designs created by Chang [46], who reinterpreted the nostalgic Taiwanese

XSL•FO
RenderX

grocery store scene. Unlike the convenience stores today, the traditional grocery stores in the 1960s sold not only rice, oil, salt, sauce, and tea, but also various colorful dried fruit preserves that filled the shelves of the entire store, making it a common memory of people in the 1960s (Figure 4). Through these traditional grocery stores' scenes, the elderly (players) felt like they entered a time corridor and were reminded of these beautiful memories during the gameplay. In addition, players were rewarded with classic movie posters after clearing the game, which they could use to share memories with the family and connect with them after the game. The rewards in this study were masterpiece movie posters hand-painted by Chang [46], who drew posters of the most popular movies in Taiwan in the 1960s (Figure 5). By presenting elderly players with the historical atmosphere in the 1960s and letting them walk through the movie scenes and the historical corridor through classic movie posters, the game also connected elderly people's memory with the story of the movie. By contrast, the remaining time interface of the game was designed by including their children as the countdown image, which symbolized their grandchildren's company during the gameplay. Therefore, the game improved the relationship between the elderly and his/her families and promoted the former to share and recall the past, thereby delaying the onset of dementia (Figure 6).

Figure 4. Traditional Taiwanese grocery stores in the 1960s.

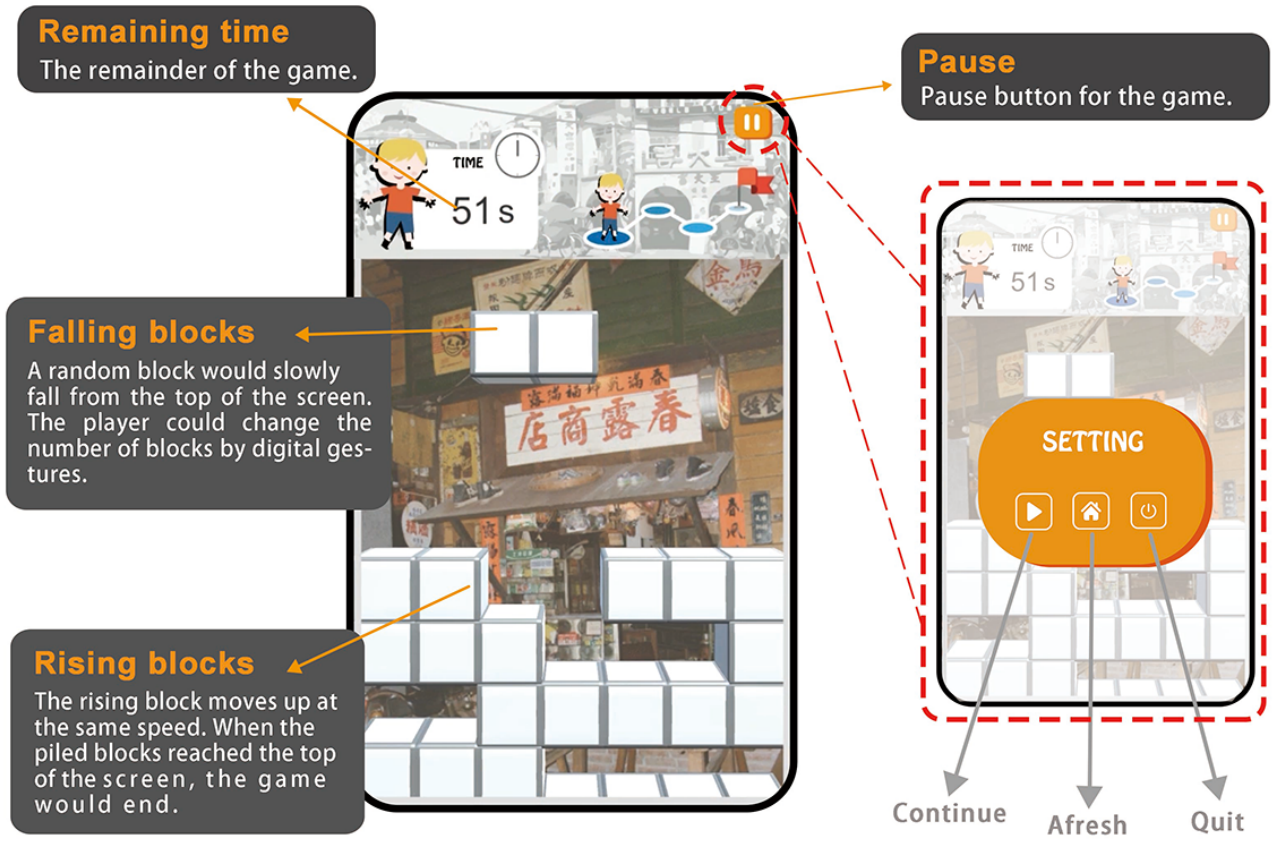


Figure 5. Movie posters in Taiwan in the 1960s designed by Chang.



XSL•FO
RenderX

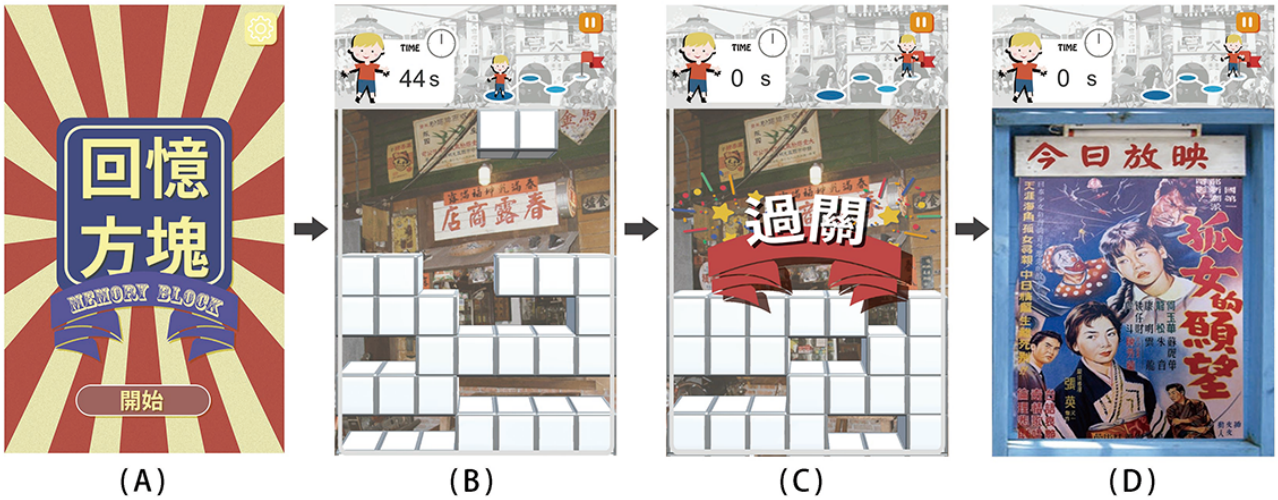
Figure 6. Description of the game interface design.



Game System Design

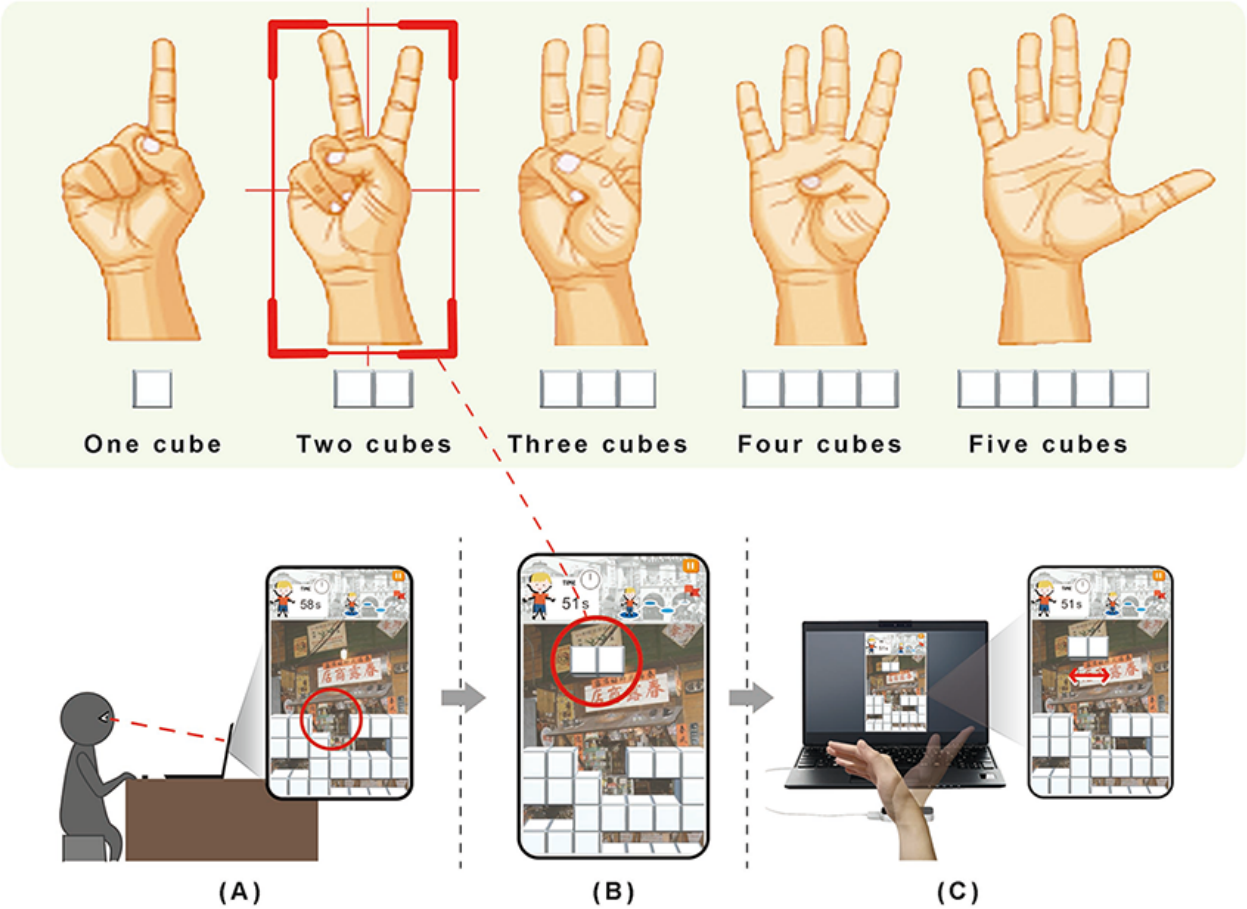
Once the interactive somatosensory game was initiated, the player was first shown the main screen (Figure 7A) before entering the game (Figure 7B). After the player cleared the level in 1 minute (Figure 7C), a bonus content would be awarded (Figure 7D). At the start of the game, a random block would slowly fall from the top of the screen (Figure 8A). During this process, the player could change the number of blocks by digital gestures (Figure 8B), followed by relocating the falling block to the desired position by moving hands left and right or accelerating the block by moving the hands down (Figure 8C). When the block reached the bottom of the screen or the top of other blocks, it would stop there, while a new random block would appear on top of the screen and start to fall. When a horizontal line was filled with blocks, it would be eliminated, increasing the player's score. The more lines eliminated simultaneously, the more points the player gets. When the piled blocks reached the top of the screen and could no longer be removed, the game would end.

Figure 7. Process of the game developed in this study.



XSL•FO
RenderX

Figure 8. Game operation. (A) Watch the number of the topmost row of blocks when a block falls; (B) change the number of randomly falling blocks through digital gestures; (C) control the movement of the block by swinging the hands left and right.



Clinical Trial of Interactive Somatosensory Game Rehabilitation

Overview

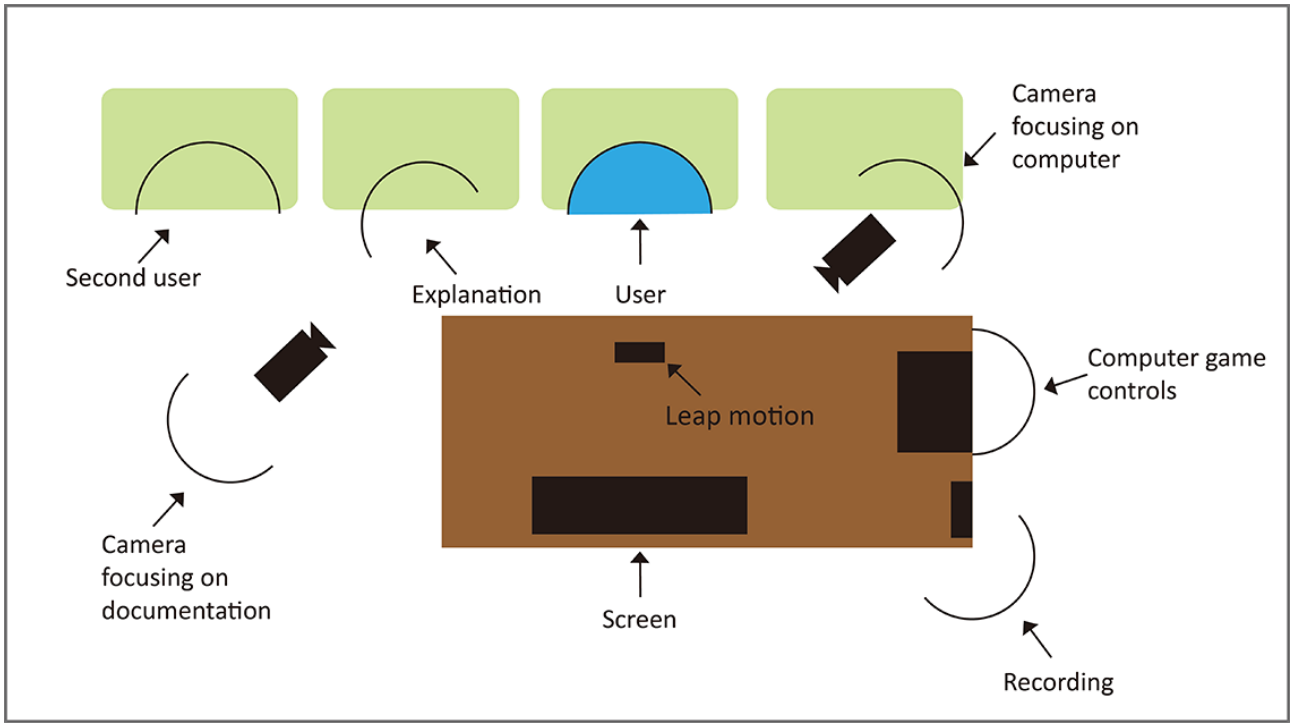
The clinical trial of the experiment was described in this section, including research patients and the experimental design (ie, TCRR and ISGR). The SUS questionnaire with a 5-point scale was adopted to collect quantitative information.

Research Patients

This study interviewed a total of 15 patients at a Clinic in Tainan, Taiwan, consisting of 7 males (5 with rehabilitation experience and 2 without) and 8 females (2 with rehabilitation experience and 6 without). The average age of the patients was 78.4 years. The inclusion criteria were elderly people over 60 years old who were diagnosed with MCI or were prone to dementia. Participants were also required to be able to cooperate with researchers, follow game instructions, and complete a subjective assessment after rehabilitation. The experimental location of this study is shown in Figure 9.

XSL•FO
RenderX

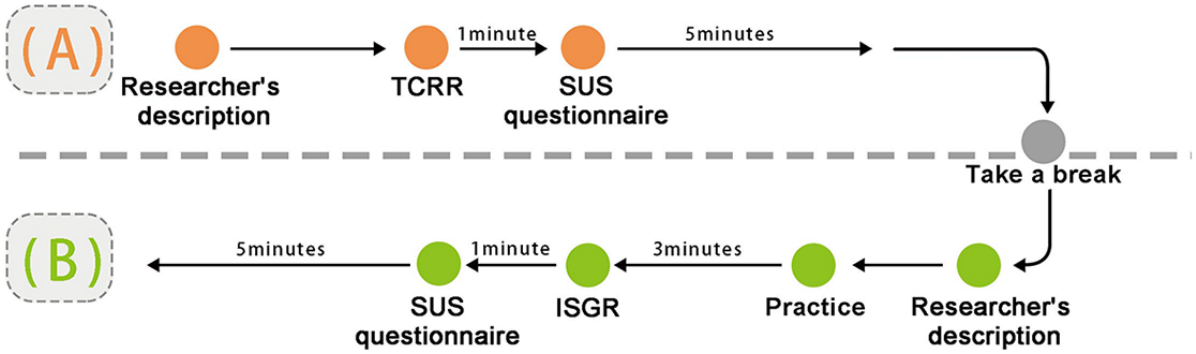
Figure 9. Experimental location.



Experimental Design

The experiment adopted a single-group design, that is, all patients would participate in both TCRR and ISGR. The questionnaire, composed of 10 questions (items), was designed based on the modified SUS [14]. Each question was scored on the 5-point scale (with 1 representing strongly disagree and 5 representing strongly agree) to understand the elderly's subjective feelings of the system during operation. The purpose of the experiment was to investigate the relationship between the use of TCRR and ISGR among the elderly and to analyze their demands (subjective feelings). The experiment was divided into 2 stages, that is, TCRR and ISGR. Out of 15 patients, 7 (5 males and 2 females) had prior rehabilitation experience, thus having operated on the TCRR before. To make the baseline of previous experience the same for all 15 patients, they all first operated on the TCRR and then on the ISGR. The experiment flowchart is shown in Figure 10.

Figure 10. Experiment flow chart. The experiment starts with a (A) TCRR, followed by (B) ISGR. ISGR: interactive somatosensory game rehabilitation; SUS: System Usability Scale; TCRR: traditional card recognition rehabilitation.



Experiment Stages

Experiment of TCRR

The TCRR used commercial poker cards for rehabilitation. During the experiment, researchers would cover the cards and place them on the table. The elders could choose any 2 cards and flip them over each time. If the 2 cards were paired (having the same pattern), the elders would gain the opportunity to continue revealing other cards, or otherwise the 2 cards would be flipped back over. The rehabilitation would end when all cards were revealed (Figure 11). In this study, after the operation method was explained by the researcher, the participant would perform the TCRR for 1 minute. Upon completion, the researcher would ask the elders about their subjective feelings of the TCRR using the 5-point scale. The contents of the SUS are listed in Table 1.

XSL•FO
RenderX

Figure 11. Traditional card recognition rehabilitation.

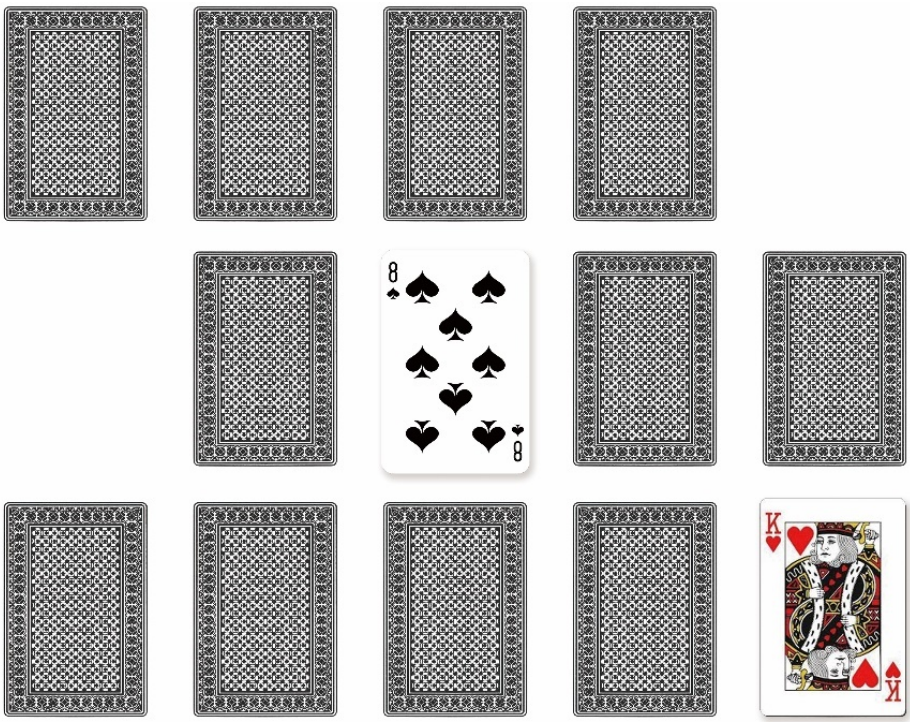


Table 1. SUS[a] questionnaire design.

Item	Question description
1	I think that I would like to use this *TCRR*[b]*/ISGR*[c] frequently.
2	I found the *TCRR/ISGR* unnecessarily complex.
3	I thought the *TCRR/ISGR* was easy to use.
4	I think that I would need the support of a technical person to be able to use this *TCRR/ISGR*.
5	I found the various functions in this *TCRR/ISGR* were well integrated.
6	I thought there was too much inconsistency in this *TCRR/ISGR*.
7	I would imagine that most people would learn to use this *TCRR/ISGR* very quickly.
8	I found the *TCRR/ISGR* very cumbersome to use.
9	I felt very confident using the *TCRR/ISGR*.
10	I needed to learn a lot of things before I could get going with this *TCRR/ISGR*.

[a]SUS: System Usability Scale.

[b]TCRR: traditional card recognition rehabilitation.

[c]ISGR: interactive somatosensory game rehabilitation.

Experiment of ISGR

Prior to the formal start of the experiment session, the researcher would first explain the operation of the game verbally, followed by 1 or 2 practices to ensure that the elderly understood the operations correctly. This process would take approximately 3 minutes. During the experiment, the game scenario was randomly selected, and the elderly would use gestures to perform the ISGR. Upon completion, the researcher would ask the elderly about his/her subjective feelings of the ISGR through the 5-point scale. The elderly would verbally answer the questions and the researcher would assist in filling out the records.

Ethics Approval

This study was approved by the National Cheng Kung University Human Research Ethics Committee (approval number NCKU HREC-F-109-497-2).

Results

Overview

The purpose of this study was to investigate whether the ISGR could improve the willingness and motivation for rehabilitation among the elderly with MCI. The study recruited 15 patients, consisting of 7 males and 8 females with an average age of 78.4 years (please see the “Methods” section). Subsequently, the

XSL•FO
RenderX

independent (unpaired) sample *t* test was adopted to analyze whether there were any differences in the willingness to use between TCRR and ISGR. This was realized by conducting 2 SUS surveys, with the first enquiring the elderly about their intention to use (willingness to use) the TCRR, and the second enquiring the same group about their intention to use (willingness to use) the ISGR. Table 2 shows the results of the 2 surveys. It should be mentioned that among the 15 patients, 8 (2 males and 6 females) had no rehabilitation experience, whereas the remaining 7 (5 males and 2 females) had prior rehabilitation experience. As Table 2 shows, the overall SUS score for the TCRR is 58.0, while for the ISGR it is 73.7 (t_{28}=–4.62, P<.001). The result shows that the ISGR has better overall usability performance (*average* +) than the TCRR (*poor*) [47,48].

To further investigate, Table 3 shows the individual item benchmarks for the TCRR and the ISGR. For the TCRR, only 2 items (out of 10 items) achieved the *average* benchmark (items 2 and 8) [47,48], thus suggesting that there was a lot of room for improvement in the usability of the TCRR. By contrast, the ISGR has 3 items reaching the *average* level (items 5, 8, and 10) and 5 reaching the *good* level (items 1-3, 7, and 9). Only items 4 and 6 need to improve further. This result indicates that the ISGR presents an average or good experience for elders with MCI when interpreting single items from the SUS [48], as compared with the TCRR. The following analyses examine the effects of prior rehabilitation experience and gender.

Table 2. The overall SUS[a] scores for the TCRR[b] and ISGR[c] (n=15 for both).

Patients	Sex	Prior rehabilitation experience: Yes or No	SUS score for TCRR[d]	SUS score for ISGR[e]
1	Male	Yes	62.5	65.0
2	Male	Yes	67.5	82.5
3	Male	No	47.5	77.5
4	Male	No	62.5	70.0
5	Male	Yes	62.5	80.0
6	Male	Yes	60.0	65.0
7	Male	Yes	57.5	75.0
8	Female	Yes	85.0	82.5
9	Female	No	47.5	60.0
10	Female	No	47.5	82.5
11	Female	No	55.0	85.0
12	Female	No	57.5	75.0
13	Female	Yes	60.0	67.5
14	Female	No	50.0	60.0
15	Female	No	47.5	77.5

[a]SUS: System Usability Scale.

[b]TCRR: traditional card recognition rehabilitation.

[c]ISGR: interactive somatosensory game rehabilitation.

[d]Mean (SD) 58.0 (10.01), so rated “poor”.

[e]Mean (SD) 73.7 (8.49), so rated “average +”.

XSL•FO
RenderX

Table 3. Basic descriptive statistics and item benchmarks for SUS[a].

Item	TCRR[b] (1=lowest, 5=highest)		TCRR score (n=15), mean (SD)	Item benchmark	ISGR[c] (1=lowest, 5=highest), mean (SD)		ISGR score (n=15), mean (SD)	Item benchmark
	No rehabilitation experience (n=8), mean (SD)	Prior rehabilitation experience (n=7), mean (SD)			No rehabilitation experience (n=8), mean (SD)	Prior rehabilitation experience (n=7), mean (SD)		
1	2.50 (0.53)	3.28 (0.75)	2.86 (0.74)	Poor –	4.25 (0.46)	4.28 (0.48)	4.27 (0.46)	Good +
2	2.87 (0.83)	2.14 (0.69)	2.53 (0.83)	Average –	2.12 (0.64)	2.00 (1.15)	2.06 (0.88)	Good –
3	3.25 (1.03)	3.42 (0.53)	3.33 (0.81)	Poor	4.37 (0.74)	4.00 (0.58)	4.20 (0.68)	Good
4	3.12 (0.99)	1.85 (0.37)	2.53 (0.99)	Poor –	2.12 (0.99)	1.85 (0.69)	2.00 (0.85)	Poor –
5	2.37 (0.51)	2.71 (1.38)	2.53 (0.99)	Poor –	3.75 (0.46)	3.71 (0.95)	3.73 (0.99)	Average +
6	2.75 (0.70)	2.14 (0.69)	2.47 (0.74)	Poor	2.87 (0.99)	2.00 (0.82)	2.46 (1.11)	Poor
7	3.12 (0.64)	3.57 (0.53)	3.33 (0.62)	Poor –	4.25 (0.71)	3.85 (0.89)	4.06 (0.80)	Good –
8	2.00 (0.53)	2.42 (0.53)	2.20 (0.56)	Average	2.00 (0.75)	2.57 (0.79)	2.27 (0.80)	Average
9	3.12 (0.83)	4.00 (0.58)	3.53 (0.83)	Poor	4.00 (0.53)	4.14 (0.69)	4.07 (0.59)	Good –
10	3.12 (1.25)	2.42 (0.79)	2.80 (1.08)	Poor –	2.12 (0.64)	2.00 (0.58)	2.07 (0.59)	Average

[a]SUS: System Usability Scale.

[b]TCRR: traditional card recognition rehabilitation.

[c]ISGR: interactive somatosensory game rehabilitation.

Analysis of the Effect of Prior Rehabilitation Experience

One objective of the study was to determine the degree of acceptance of the novel ISGR among the elderly who previously went through TCRR, and to further understand whether prior rehabilitation experience would affect the willingness to use the novel rehabilitation method. Therefore, the 2 rehabilitation methods were analyzed separately for those with and without rehabilitation experience, the results of which are shown in Table 4. Table 4 revealed that regardless of prior rehabilitation experience, there were no significant differences in the answers of the elderly after TCRR or ISGR. For example, for item 1 in the SUS, "I think that I would like to use this TCRR/ISGR frequently," the *t* value of TCRR and ISGR was –0.64 (*df*=13, *P*=.53) and –0.15 (*df*=13, *P*=.89), respectively, both of which indicated insignificant differences. This result showed that for elders, their willingness to use the ISGR was not significantly affected by prior rehabilitation experience. Hence, regardless of prior rehabilitation experience, the elderly demonstrated consistent subjective feelings of use (willingness to use) toward both TCRR and ISGR, that is, their intention to use and motivation for ISGR would not be compromised because of complicated operations. This finding was further supported by the elders' answers to item 3, "I thought the TCRR/ISGR was easy to use." The *t* value of TCRR and ISGR for this question was –0.41 (*df*=13, *P*=.69) and 1.08 (*df*=13, *P*=.30), respectively, again showing insignificant differences. Further analysis suggested that despite insignificant differences, the average score of ISGR (4.20) was considerably higher than that of TCRR (3.33). This result revealed the fact that the elderly believed the ISGR was easier to operate/use (please refer to the following sections for more details).

XSL•FO
RenderX

Table 4. Independent sample *t* test of the effect of prior rehabilitation experience.

Item	TCRR[a]		ISGR[b]	
	t_{13} value	*P* value	t_{13} value	*P* value
1	–0.64	.53	–0.15	.89
2	1.83	.09	0.26	.79
3	–0.41	.69	1.08	.30
4	2.03	.06	0.60	.56
5	–0.14	.89	0.09	.93
6	1.68	.12	1.85	.09
7	–1.45	.17	0.95	.36
8	–0.90	.38	–0.96	.36
9	–1.10	.29	–0.45	.66
10	1.27	.23	0.39	.70

[a]TCRR: traditional card recognition rehabilitation.

[b]ISGR: interactive somatosensory game rehabilitation.

Analysis of the Effect of Gender

In this section, the 2 rehabilitation methods were analyzed separately for patients of different gender (ie, males and females) to see if the gender factor exhibited any effects on the willingness to use the TCRR or ISGR. Table 5 indicates that participants of different gender showed significant differences in the "Easy to learn" section of the ISGR. For example, for item 7, "I would imagine that most people would learn to use this TCRR/ISGR very quickly," the *t* value of TCRR was –0.54 (*df*=13, *P*=.59, indicating insignificant difference between gender), whereas that of ISGR was 2.71 (*df*=13, *P*=.02), which showed significant differences. The results suggested that the novel ISGR developed in this study was easier to learn for females. Based on the findings, it was concluded that while the willingness to use (item 1: t_{13}=0.98, *P*=.34, insignificant difference) and ease of use (item 3: t_{13}=1.08, *P*=.30, insignificant difference) of the ISGR were not affected by the gender factor, this was not the case for the ease to learning (item 7). The results indicated that females were more willing to learn new things than males were; this should be investigated in future research.

Table 5. Independent sample *t* test of the effect of gender.

Item	TCRR[a]		ISGR[b]	
	t_{13} value	*P* value	t_{13} value	*P* value
1	0.04	.96	0.98	.34
2	1.08	.30	0.85	.41
3	–0.41	.69	1.08	.30
4	0.90	.38	0.60	.56
5	0.90	.38	–1.42	.18
6	0.18	.86	0.65	.53
7	–0.54	.59	2.71	.02[c]
8	–0.54	.60	–0.08	.93
9	–0.77	.45	0.39	.70
10	1.27	.23	0.39	.70

[a]TCRR: traditional card recognition rehabilitation.

[b]ISGR: interactive somatosensory game rehabilitation.

[c]Significant value.

XSL•FO
RenderX

Comparison Between TCRR and ISGR

In the previous sections, effects of "prior rehabilitation experience" and "gender" were analyzed. The results showed that prior rehabilitation experience did not affect the elderly's willingness to use (item 1) and ease of use (item 3). However, the ease of learning (item 7) was affected by the gender factor. In this section, elders with and without rehabilitation experience were divided into separate groups to analyze their respective subjective feelings (willingness to use) of the 2 rehabilitation methods (ie, TCRR and ISGR; as shown in Table 6).

From Table 6 (column 2), significant differences between the 2 rehabilitation methods were observed on item 1 (t_{28}=–8.27, P<.001), item 3 (t_{28}=–3.17, P<.001), item 5, "I found the various functions in this TCRR/ISGR were well integrated" (t_{28}=–5.07, P<.001), item 7 (t_{28}=–2.81, P=.01), and item 10, "I needed to learn a lot of things before I could get going with this TCRR/ISGR" (t_{28}=2.04, P=.03). The results suggested that the novel ISGR proposed in this study showed better "willingness to use," "ease of use," "integration," "ease of learning," and "system complexity" than TCRR. The elders indicated that they were more willing to use the novel method continuously, which was of great significance for the long-term rehabilitation of patients with dementia.

Table 6. Independent sample *t* test for TCRR[a] and ISGR[b].

Item	Comparison of the 2 rehabilitation methods (n=15)		Prior rehab experience (n=7)		No rehab experience (n=8)	
	t_{28} value	*P* value	t_{12} value	*P* value	t_{14} value	*P* value
1	–8.27	<.001[c]	–4.62	<.001[c]	–7.00	<.001[c]
2	1.49	.15	0.28	.78	2.02	.06
3	–3.17	<.001[c]	–1.92	.08	–2.50	.02[c]
4	0.72	.48	0.00	.99	0.85	.41
5	–5.07	<.001[c]	–2.50	.03[c]	–5.60	<.001[c]
6	0.00	.99	0.35	.11	–0.30	.75
7	–2.81	.01[c]	–0.72	.48	–3.33	<.001[c]
8	0.39	.70	0.28	.79	0.27	.79
9	–1.21	.23	–0.42	.68	–1.18	.26
10	2.04	.03[c]	1.16	.27	2.02	.06

[a]TCRR: traditional card recognition rehabilitation.

[b]ISGR: interactive somatosensory game rehabilitation.

[c]Significant value.

Further analysis was performed by dividing participants into those with and without prior rehabilitation experience. Table 6 (columns 3 and 4) showed that for the elderly with and without prior rehabilitation experience toward TCRR and ISGR, significant differences were seen in item 1 on "willingness to use" (t_{12}=–4.26, P<.001 and t_{14}=–7.00, P<.001, respectively) and item 5 on "integration" (t_{12}=–2.50, P=.03 and t_{14}=–5.60, P<.001, respectively). This is consistent with the previous finding that "Regardless of prior rehabilitation experience, the elderly consistently showed improved willingness to use the novel ISGR over the TCRR and believed that the integration of ISGR was better."

Comparison between the 2 rehabilitation methods suggested that for the elderly without prior rehabilitation experience (Table 6, column 4), significant differences were seen in item 3 on "ease of use" (t_{14}=–2.50, P=.20) and item 7 on "ease of learning" (t_{14}=–3.33, P<.001). This finding indicated that in terms of "ease of use" and "ease of learning," the ISGR was easier for the elderly, especially for those without prior rehabilitation experience to master and operate. Thus, if an individual was exposed to the 2 rehabilitation methods for the first time, he or she would find the ISGR much easier to learn and use, which would in turn significantly affect the willingness of the elderly to continue rehabilitation. The results also echoed the conclusions of previous sections.

For item 10 on "system complexity" (I need to learn a lot...), the 2 rehabilitation methods showed significant differences (t_{28}=2.04, P=.03; Table 5, column 2), but differences between the elderly with and without prior rehabilitation experience were insignificant (t_{12}=1.16, P=.27 and t_{14}=2.02, P=.06, respectively). This result suggested that regardless of prior rehabilitation experience, the elderly believed that it took them a decent amount of time to learn and become familiar with the TCRR before they were willing to continue rehabilitation. By contrast, the elderly could rapidly master the novel ISGR (because they could use intuitive gesture recognition during rehabilitation) and were more willing to use it in the long term.

XSL•FO
RenderX

Discussion

Principal Findings

In summary, elderly participants provided positive feedback on the ISGR developed in this study. The novel method would not cause the elderly to feel stressful or reluctant under the same activity and treatment arrangement of traditional methods, making it applicable for the general elderly population. Furthermore, the combination of rehabilitation with interactive games not only improved the user's motivation for rehabilitation, but also enhanced the ease of use and learning of the system through intuitive gesture recognition technology, thereby promoting the user's willingness to continuously engage in rehabilitation.

Recommendations for Future Research

This study proposed the integration of a novel interactive somatosensory game into the hand rehabilitation of the elderly with MCI. We also provide suggestions on "interactive somatosensory" and "game design" for future research.

Interactive Somatosensory

During literature review, questionnaire survey, and experiments, it was found that patients with dementia were highly interested in interactive somatosensory games. Furthermore, the replacement of traditional rehabilitation methods with intuitive gesture recognition made it easy for the elderly to learn to use rehabilitation devices, which was consistent with findings of previous studies [5,27]. Therefore, it is recommended that subsequent research can develop various cognitive training modules based on the ISGR method developed in this study. Furthermore, a system on the medical network that allows medical institutions of different levels and even home caregivers to download interactive somatosensory games can be established so that the rehabilitation of the elderly can be performed through telemedicine or homecare [24-26].

Game Design

Observations during experiments indicated that using a nostalgic style in the game design was more effective in reminding the elderly of memories and prompting them to share their experience, which was consistent with findings of previous research [36,37]. Furthermore, interactive games not only demonstrate a cognitive training effect, but are also capable of increasing user stickiness by introducing various game scenarios and difficulty levels, thereby productively improving the elderly's willingness and motivation for rehabilitation. Therefore, when designing interactive games, researchers should deeply discuss the humane, social, cultural, and other features of the elderly population and integrate them into the game content to make the elderly feel familiar with the nostalgic scenes in the game. This finding is consistent with the results of the study by Cotelli et al [38] and should be further exploited in future research.

Conclusions

With an aging society as well as an increasing number of patients with dementia, it is of great significance to prevent and delay the incidence of dementia. Exercise can not only effectively activate the brain, but also strengthen the effect of brain activities through hand exercise as well as improving the blood circulation of the body, making it beneficial in preventing and delaying dementia. This study combined hand exercise with nostalgic therapy and integrated interactive games into the cognitive training of the elderly, so as to increase interest in rehabilitation and improve the elderly's motivation to engage in rehabilitation in the long term. By asking the elderly to participate in both TCRR and ISGR, this study investigated the willingness of use of the elderly toward these 2 rehabilitation methods. Furthermore, effects of gender and prior rehabilitation experience on these 2 rehabilitation methods were explored to understand the elderly's thoughts and feelings of ISGR. Results showed that the ease of learning of the 2 methods was considerably different. For elderly people without prior rehabilitation experience, the ISGR was easier to learn and operate than the TCRR, and females performed better in learning the novel method. The results suggested that the novel ISGR developed in this study could not only improve the elderly's motivation for rehabilitation, but also promote them to continuously participate in rehabilitation. Therefore, it can be widely applied to other rehabilitation fields (such as home rehabilitation or telemedicine) in the future.

Acknowledgments

This research was supported in part by the Ministry of Science and Technology, Taiwan under Grant Nos. MOST 108-2628-E-006-007-MY3 and MOST 111-2923-E-006-002-MY3. We are grateful to the anonymous reviewers for their valuable comments and suggestions.

Conflicts of Interest

None declared.

References

1. World Health Organization. Dementia. World Health Organization. 2021 Sep 02. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2020-01-27]
2. Ministry of Health and Welfare. The Guideline of Dementia. Ministry of Health and Welfare. 2017 Dec 19. URL: https://1966.gov.tw/LTC/cp-4020-42469-201.html [accessed 2020-01-27]

XSL•FO
RenderX

3. Song Y, Wang J. Overview of Chinese research on senile dementia in mainland China. Ageing Res Rev 2010 Nov;9 Suppl 1:S6-12. [doi: 10.1016/j.arr.2010.08.007] [Medline: 20813206]
4. Lai R, Tensil M, Kurz A, Lautenschlager NT, Diehl-Schmid J. Perceived Need and Acceptability of an App to Support Activities of Daily Living in People With Cognitive Impairment and Their Carers: Pilot Survey Study. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16928 [FREE Full text] [doi: 10.2196/16928] [Medline: 32735223]
5. Fried I, Mukamel R, Kreiman G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. Neuron 2011 Feb 10;69(3):548-562 [FREE Full text] [doi: 10.1016/j.neuron.2010.11.045] [Medline: 21315264]
6. Moizer J, Lean J, Dell'Aquila E, Walsh P, Keary A, O'Byrne D, et al. An approach to evaluating the user experience of serious games. Computers & Education 2019 Jul;136:141-151. [doi: 10.1016/j.compedu.2019.04.006]
7. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
8. Laamarti F, Eid M, El Saddik A. An Overview of Serious Games. International Journal of Computer Games Technology 2014;2014:1-15. [doi: 10.1155/2014/358152]
9. Kleschnitzki JM, Grossmann I, Beyer R, Beyer L. Modification in the Motor Skills of Seniors in Care Homes Using Serious Games and the Impact of COVID-19: Field Study. JMIR Serious Games 2022 May 10;10(2):e36768 [FREE Full text] [doi: 10.2196/36768] [Medline: 35536610]
10. Held JPO, Yu K, Pyles C, Veerbeek JM, Bork F, Heining S, et al. Augmented Reality-Based Rehabilitation of Gait Impairments: Case Report. JMIR Mhealth Uhealth 2020 May 26;8(5):e17804 [FREE Full text] [doi: 10.2196/17804] [Medline: 32452815]
11. Ling Y, Ter Meer LP, Yumak Z, Veltkamp RC. Usability Test of Exercise Games Designed for Rehabilitation of Elderly Patients After Hip Replacement Surgery: Pilot Study. JMIR Serious Games 2017 Oct 12;5(4):e19 [FREE Full text] [doi: 10.2196/games.7969] [Medline: 29025696]
12. Allegue DR, Kairy D, Higgins J, Archambault PS, Michaud F, Miller WC, et al. A Personalized Home-Based Rehabilitation Program Using Exergames Combined With a Telerehabilitation App in a Chronic Stroke Survivor: Mixed Methods Case Study. JMIR Serious Games 2021 Aug 31;9(3):e26153 [FREE Full text] [doi: 10.2196/26153] [Medline: 34132649]
13. Zarit SH, Chiusano C, Harrison AS, Sewell L, Krause C, Liu Y. Rehabilitation of persons with dementia: using technology to improve participation. Aging Ment Health 2021 Mar 16;25(3):543-550. [doi: 10.1080/13607863.2020.1711864] [Medline: 31941356]
14. Brooke J. A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London, UK: Taylor and Francis; 1996:189-194.
15. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
16. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
17. Payne J. Integrating Application Security into Software Development. IT Prof 2010 Mar;12(2):6-9. [doi: 10.1109/mitp.2010.58]
18. Lunardini F, Luperto M, Romeo M, Basilico N, Daniele K, Azzolino D, et al. Supervised Digital Neuropsychological Tests for Cognitive Decline in Older Adults: Usability and Clinical Validity Study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17963 [FREE Full text] [doi: 10.2196/17963] [Medline: 32955442]
19. Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment. Neurology 2017 Dec 27;90(3):126-135. [doi: 10.1212/wnl.0000000000004826]
20. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999 Mar 01;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
21. Harvey L, Mitchell R, Brodaty H, Draper B, Close J. Differing trends in fall-related fracture and non-fracture injuries in older people with and without dementia. Arch Gerontol Geriatr 2016 Nov;67:61-67. [doi: 10.1016/j.archger.2016.06.014] [Medline: 27434743]
22. Yang ZF. Fall is the beginning of health deterioration. Rainbow Newsletter 2015;10:1.
23. Jia R, Liang J, Xu Y, Wang Y. Effects of physical activity and exercise on the cognitive function of patients with Alzheimer disease: a meta-analysis. BMC Geriatr 2019 Jul 02;19(1):181 [FREE Full text] [doi: 10.1186/s12877-019-1175-2] [Medline: 31266451]
24. Larson EB, Wang L, Bowen JD, McCormick WC, Teri L, Crane P, et al. Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. Ann Intern Med 2006 Jan 17;144(2):73-81 [FREE Full text] [doi: 10.7326/0003-4819-144-2-200601170-00004] [Medline: 16418406]
25. Li X, Zhang T. Healthy Education Exercise Prevention and Treatment Prescription for Senile Dementia. Eurasia J Math Sci Tech Edu 2017;13(12):8051-8055. [doi: 10.12973/ejmste/80763]
26. van Praag H, Fleshner M, Schwartz MW, Mattson MP. Exercise, Energy Intake, Glucose Homeostasis, and the Brain. Journal of Neuroscience 2014 Nov 12;34(46):15139-15149. [doi: 10.1523/jneurosci.2814-14.2014]

XSL•FO
RenderX

27. Voss MW, Soto C, Yoo S, Sodoma M, Vivar C, van Praag H. Exercise and Hippocampal Memory Systems. Trends Cogn Sci 2019 Apr;23(4):318-333 [FREE Full text] [doi: 10.1016/j.tics.2019.01.006] [Medline: 30777641]
28. di Pellegrino G, Làdavas E. Peripersonal space in the brain. Neuropsychologia 2015 Jan;66:126-133. [doi: 10.1016/j.neuropsychologia.2014.11.011] [Medline: 25448862]
29. Gardini S, Venneri A, McGeown WJ, Toraci C, Nocetti L, Porro CA, et al. Brain Activation Patterns Characterizing Different Phases of Motor Action: Execution, Choice and Ideation. Brain Topogr 2016 Sep 12;29(5):679-692. [doi: 10.1007/s10548-016-0491-5] [Medline: 27072014]
30. Chen S. Fascinating Dermatoglyphics. Taipei, Taiwan: New Sprouts Publishers Inc; Sep 01, 2004.
31. Hasegawa Y. Stimulating the Thumbs to Make the Brain Younger. Taipei, Taiwan: Yuan-Liou Publishing Co; 2017.
32. Garrido Navarro JE, Ruiz Penichet VM, Lozano Pérez MD. Movement-based interaction applied to physical rehabilitation therapies. J Med Internet Res 2014 Dec 09;16(12):e281 [FREE Full text] [doi: 10.2196/jmir.3154] [Medline: 25491148]
33. Da Gama A, Fallavollita P, Teichrieb V, Navab N. Motor Rehabilitation Using Kinect: A Systematic Review. Games for Health Journal 2015 Apr;4(2):123-135. [doi: 10.1089/g4h.2014.0047] [Medline: 26181806]
34. Smeragliuolo AH, Hill NJ, Disla L, Putrino D. Validation of the Leap Motion Controller using markered motion capture technology. J Biomech 2016 Jun 14;49(9):1742-1750. [doi: 10.1016/j.jbiomech.2016.04.006] [Medline: 27102160]
35. Leong K, Chen Y, Chen J. Dementia: a focused review (in Chinese). Journal of Internal Medicine of Taiwan 2014;25(3):151-157. [doi: 10.1093/med/9780199358274.003.0091]
36. Satorres E, Viguer P, Fortuna F, Meléndez JC. Effectiveness of instrumental reminiscence intervention on improving coping in healthy older adults. Stress Health 2018 Apr 18;34(2):227-234. [doi: 10.1002/smi.2776] [Medline: 28834143]
37. Cuevas PEG, Davidson PM, Mejilla JL, Rodney TW. Reminiscence therapy for older adults with Alzheimer's disease: A literature review. Int J Ment Health Nurs 2020 Jun 26;29(3):364-371. [doi: 10.1111/inm.12692] [Medline: 31984570]
38. Cotelli M, Manenti R, Zanetti O. Reminiscence therapy in dementia: a review. Maturitas 2012 Jul;72(3):203-205. [doi: 10.1016/j.maturitas.2012.04.008] [Medline: 22607813]
39. Kim H, Chen JS. The Memorable Travel Experience and Its Reminiscence Functions. Journal of Travel Research 2018 May 15;58(4):637-649. [doi: 10.1177/0047287518772366]
40. Melendez JC, Torres M, Redondo R, Mayordomo T, Sales A. Effectiveness of follow-up reminiscence therapy on autobiographical memory in pathological ageing. Int J Psychol 2017 Aug 17;52(4):283-290. [doi: 10.1002/ijop.12217] [Medline: 26377350]
41. Woods B, O'Philbin L, Farrell E, Spector A, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 01;3:CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
42. Duru Aşiret G, Kapucu S. The Effect of Reminiscence Therapy on Cognition, Depression, and Activities of Daily Living for Patients With Alzheimer Disease. J Geriatr Psychiatry Neurol 2016 Jan 06;29(1):31-37. [doi: 10.1177/0891988715598233] [Medline: 26251112]
43. Vokorokos L, Mihal'ov J, Chovancová E. Motion sensors: Gesticulation efficiency across multiple platforms. 2016 Presented at: IEEE 20th Jubilee International conference on Intelligent Engineering Systems; June 30-July 02, 2016; Budapest, Hungary p. 293-298. [doi: 10.1109/ines.2016.7555139]
44. Guzsvinecz T, Szucs V, Sik-Lanyi C. Suitability of the Kinect Sensor and Leap Motion Controller-A Literature Review. Sensors (Basel) 2019 Mar 02;19(5):1072 [FREE Full text] [doi: 10.3390/s19051072] [Medline: 30832385]
45. Lu W, Tong Z, Chu J. Dynamic Hand Gesture Recognition With Leap Motion Controller. IEEE Signal Process. Lett 2016 Sep;23(9):1188-1192. [doi: 10.1109/lsp.2016.2590470]
46. DAVID CHANG. 50th Museum. 2011 Mar 25. URL: https://blog.xuite.net/colastore1688140/twblog/127526648 [accessed 2020-02-10]
47. Sauro J. Interpreting single items from the SUS. Measuring U. 2018 Jul 11. URL: https://measuringu.com/sus-items [accessed 2022-04-28]
48. Lewis J, Sauro J. Item benchmarks for the system usability scale. Journal of Usability Studies 2018;13(3):158-167. [doi: 10.5555/3294033.3294037]

Abbreviations

ISGR: interactive somatosensory game rehabilitation
LM: leap motion
MCI: mild cognitive impairment
SUS: System Usability Scale
TCRR: traditional card recognition rehabilitation
WHO: World Health Organization

XSL•FO
RenderX

Edited by N Zary; submitted 03.04.22; peer-reviewed by J Brooke, N Chalghaf; comments to author 25.04.22; revised version received 22.05.22; accepted 12.06.22; published 14.07.22.

Please cite as:
Chang CH, Yeh CH, Chang CC, Lin YC
Interactive Somatosensory Games in Rehabilitation Training for Older Adults With Mild Cognitive Impairment: Usability Study
JMIR Serious Games 2022;10(3):e38465
URL: https://games.jmir.org/2022/3/e38465
doi:10.2196/38465
PMID:35834303

©Chien-Hsiang Chang, Chung-Hsing Yeh, Chien-Cheng Chang, Yang-Cheng Lin. Originally published in JMIR Serious Games (https://games.jmir.org), 14.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality as a Therapy Tool for Walking Activities in Pediatric Neurorehabilitation: Usability and User Experience Evaluation

Corinne Ammann-Reiffer[1,2], PT, PhD; Andrina Kläy[1,2], MSc; Urs Keller[1,2], PhD

[1]Research Department, Swiss Children's Rehab, University Children's Hospital Zurich - Eleonore Foundation, Affoltern am Albis, Switzerland
[2]Children's Research Center, University Children's Hospital Zurich - Eleonore Foundation, Zurich, Switzerland

Corresponding Author:
Corinne Ammann-Reiffer, PT, PhD
Research Department
Swiss Children's Rehab
University Children's Hospital Zurich - Eleonore Foundation
Mühlebergstrasse 104
Affoltern am Albis, 8910
Switzerland
Phone: 41 447625297
Email: corinne.ammann@kispi.uzh.ch

Abstract

Background: Many essential walking activities in daily life, such as crossing a street, are challenging to practice in conventional therapeutic settings. Virtual environments (VEs) delivered through a virtual reality (VR) head-mounted display (HMD) would allow training such activities in a safe and attractive environment. Furthermore, the game-like character and high degree of immersion in these applications might help maintain or increase children's motivation and active participation during the rehabilitation process.

Objective: This study aimed to investigate the usability, user experience, and acceptability of an immersive VE experienced through a VR HMD to train everyday life walking activities in pediatric neurorehabilitation.

Methods: In a cross-sectional study, 21 youths (median age 12.1 years; range 6.8-17.7 years) with a neuromotor impairment undergoing inpatient or outpatient neurorehabilitation tested a VE experienced through the VR HMD Oculus Quest. The participants, accompanied by their physiotherapists, moved freely around a 4.4 by 10-meter VE, displaying a magical forest and featuring various gamified everyday activities in different game designs. Using their hands, represented in the VE, the participants could interact with the virtual objects placed throughout the VE and trigger visual and auditory feedback. Symptoms of cybersickness were checked, and usability, user experience, and acceptability were evaluated using customized questionnaires with a visual analog scale for youths and a 5-point Likert scale for their therapists.

Results: None of the participants reported any signs of cybersickness after 20 minutes of VR HMD exposure time. They rated comfort (median 10/10) and movement ability (median 10/10) with the VR HMD as high. The VE was perceived as being really there by the majority (median 8/10), and the participants had a strong feeling of spatial presence in the VE (median 9.5/10). They enjoyed exploring the virtual world (median 10/10) and liked this new therapy approach (median 10/10). Therapists' acceptance of the VR HMD was high (4/5). There were 5 patients that needed more support than usual, mainly for supervision, when moving around with the VR HMD. Otherwise, therapists felt that the VR HMD hardly affected their patients' movement behavior (median 4.75/5), whereas it seemed to increase their level of therapy engagement (median 4/5) compared to conventional physiotherapy sessions.

Conclusions: This study demonstrates the usability of an immersive VE delivered through a VR HMD to engage youths in the training of everyday walking activities. The participants' and therapists' positive ratings on user experience and acceptance further support the promising application of this technology as a future therapeutic tool in pediatric neurorehabilitation.

(JMIR Serious Games 2022;10(3):e38509) doi:10.2196/38509

KEYWORDS

rehabilitation; pediatric; child; adolescent; walking; feasibility study; virtual reality; head-mounted display; therapy; tool; user; usability; walking; visual; auditory; feedback; youth

XSL•FO
RenderX

Introduction

Pediatric neurorehabilitation strives to provide patients with the greatest possible degree of independence in everyday life [1]. Therapies aimed at increasing children's independence require them to actively work on their limitations and push their physical limits, which takes weeks, months, or even years—depending on the nature and severity of their impairment. Active involvement, perseverance, and adherence to the therapy program are crucial ingredients for these therapies, which incorporate various motor learning principles, to be successful [2]. However, it is particularly challenging for children to maintain all these qualities throughout a lengthy rehabilitation stay.

In recent years, computer technology applications that create virtual environments (VEs) have emerged more and more in the field of rehabilitation [3]. By merging the physical and virtual worlds, VEs enable task-specific training and can provide an ecologically valid environment similar to the real world [4]. Thereby, they offer new options for therapies and outcome assessment [5] while ensuring the safety of the therapy setting [4]. In addition, VEs can help increase children's motivation and therapy adherence during the rehabilitation process, which may result in more training repetitions [6,7]. Further advantages are the enriched environments, exercise gamification, possibility of task-specific training incorporating variations and real-time feedback on performance, and easily adjustable difficulty levels to account for children's different motor abilities [4,8,9]. When compared to conventional therapy or controls, VE interventions on the upper extremities, postural control, and balance in children with cerebral palsy showed a strong effect in improving motor functions [8]. So far, VE interventions targeting gait function in children with neuromotor impairment were considerably less frequent, and consequently, evidence on their effectiveness remains limited [6-8,10]. However, evidence derived from adult populations is promising. VE interventions effectively improved balance, gait functions, and mobility in various groups of patients with neurological disorders [6,11,12].

VEs differ regarding the display device, level of immersion, and type of interaction and can be delivered by custom-built systems as well as affordable off-the-shelf options [4,8]. To date, gait interventions using VEs mainly rely on nonimmersive flat-screen VEs. Although these approaches increase the enjoyment, motivation, and adherence toward the therapy program in pediatric and adult patients [6,7], they offer only limited possibilities in terms of interaction, sensorimotor contingencies, and illusions [13]. In contrast to these 2D VEs, virtual reality (VR) head-mounted displays (HMDs) provide a stereoscopic 3D view in a completely simulated environment while blocking the views of the real surroundings and user's body. As some of these devices are wireless, they offer optimal prerequisites for motivating and immersive training of walking activities [13].

VR HMDs have already been shown to be feasible for balance or gait training in the older population [14] or adult patients with a neurological disorder [15,16]. However, in pediatric patients, the evidence of the VR HMDs' feasibility and acceptability is very limited and restricted to their use in static positions, mainly during sitting or lying [17-20]. We have previously examined and reported on the usability and acceptability of 2 different HMDs (1 VR HMD and 1 mixed reality HMD) in children undergoing inpatient neurorehabilitation [21]. Although our first results were promising regarding an application of mixed reality or VR HMDs while moving around, the wearing time was short, and the VE did not allow for interactions. In this study, we now aimed to investigate the usability, user experience, and acceptability of an immersive VE with different game applications experienced through a VR HMD to train everyday life walking activities in children and adolescents with neuromotor impairments.

Methods

Participants

Children and adolescents aged 6 to 18 years with a neuromotor impairment and undergoing inpatient or outpatient rehabilitation at the Swiss Children's Rehab (SCR) of the University Children's Hospital Zurich were eligible. We aimed to have a diverse group of participants in terms of age, diagnosis, mobility level, visual acuity level, and cognitive abilities to test the VR HMD in a heterogeneous group representing the patient composition at the SCR. Exclusion criteria were inability to follow verbal instructions, uncorrectable severe visual impairment, and a history of seizures or taking anticonvulsant medication. Written informed consent and assent was obtained from the legal representatives and participating children and adolescents.

VE Specifications

The VE, representing a magical forest (Figure 1), was created in the game engine Unity (version 2019.4.6; Unity Technologies). We chose the forest as the VE because it allowed integrating many everyday walking tasks considered important by parents and adolescents with neuromotor disorders [22]. Further, a forest represents a peaceful environment that allows the children to concentrate on tasks without being too distracted. Additionally, it can be appealing regardless of the user's age and represents a situation that is not easily accessible in real life for many patients due to their functional impairments. Walking tasks that can be performed in the VE involved stepping over various obstacles such as roots, a tree log, stones, or a small creek; opening and closing the door to a hut; stepping over the door sill and moving around in the hut's confined space; and picking sweets from a tree and carrying them to a badger to feed it. Further, the VE featured various mushrooms to play drums, lanterns to carry around, and a small tent that the children and adolescents could crawl into to grab a chocolate bar. The interactive objects should draw the participants' focus to the inner walkable workspace and encourage them to walk around, whereas the normal objects such as hills and trees at the periphery should not attract their attention.

The participants experienced the VE through a commercially available VR HMD, the Oculus Quest 1 (Facebook Technologies). The Oculus Quest is a stand-alone device with 6 degrees of freedom and 4 integrated cameras, which enable

XSL•FO
RenderX

room and hand tracking. The chosen visual design considered the performance of the Oculus Quest and the fact that a cartoon-like style is often appealing to children. The participants could freely move around the 4.4 by 10-meter VE (determined by the available space of our room and maximal given play zone of the Oculus Quest). We did not use the standard Oculus controllers as input devices, as holding or operating the controllers may have been difficult for some of our patients due to limited finger motor skills or reliance on a mobility aid. Therefore, we used the Oculus hand tracking that allows the participants to use their hands, which are displayed in a realistic and size-adapted manner in the VE, intuitively for interactions with the virtual objects. Possibilities for such interactions were placed throughout the VE and triggered auditory and visual feedback to motivate the participants to move around as much as possible.

The magical forest VE featured a *free exploration mode*, in which the children and adolescents could move around without specific instructions, and 3 different game conditions. In the *orientation game*, a picture frame revealed, upon being touched, an image of the location of the next picture frame, which the participant then had to find and touch, etc. In the *apple game*, the participants had to collect apples spread all over the VE and bring them to a basket in the hut. In the *scoring game*, the task was to score as many points as possible by interacting with the objects. Between interactions, participants had to cover at least 2 meters (indicated by a green bar at the top of the field of view) before they were able to score their next point.

A video of pediatric neurorehabilitation patients experiencing and interacting with the immersive virtual environment is shown in Multimedia Appendix 1.

Figure 1. Virtual environment with (A) a magic forest, (B) the orientation game, (C) the apple game, and (D) the scoring game.



Ethical Considerations

The ethics committee of the Canton Zurich confirmed through a clarification of responsibility that approval for this cross-sectional study, which took place in the gait laboratory of the SCR, was not needed (Req-2020-00757).

Protocol

After the participants were informed about the test procedure, the VR HMD was adjusted to their heads, and its optimal position was checked. The participants then started with the *free exploration mode*, which lasted 5 minutes. This mode allowed them to get used to the VR HMD and familiarize themselves with the VE and the interaction possibilities.

XSL•FO
RenderX

Subsequently, the participants took off the VR HMD and answered short questions regarding potential symptoms of cybersickness, using the Virtual Reality Sickness Questionnaire (VRSQ) [23]. Thereafter, they played the 3 games described above, each lasting 5 minutes, in randomized order. Patients could play the 3 games without pausing or take off the VR HMD for a short break in between. Following the 3 games, the participants again answered the VRSQ and a short set of questions. A physiotherapist accompanied the patients to provide assistance if needed and observe their movement behavior and answered a questionnaire at the end of the session.

Outcome Measures

Participant characteristics were retrieved from the patients' medical records. Their functional level of mobility was rated by the physiotherapists with the Gillette Functional Assessment Questionnaire walking scale (FAQ) and the Functional Mobility Scale (FMS) [24]. The FAQ quantifies a range of walking abilities in daily life on an ordinal scale from 1 to 10, whereas the FMS complements the information by assessing the assistive device used over 5, 50, and 500 meters on an ordinal scale from 1 to 6.

Besides the VRSQ, the investigator used a customized questionnaire covering the aspects of comfort, fun, presence in the VE, and immersion to ask the participants about their experience with the VE and VR HMD. The questions mainly consisted of a subset of items from the Comfort Rating Scale [25], the Igroup Presence Questionnaire [26], and the Presence questionnaire 2.0 [27]. Further, the participants were asked how strongly they felt that the objects and environment they saw were really there. Our understanding was that if the environment and objects were experienced as actually being there, it would indicate an increased immersion and user experience. In turn, a positive user experience could result in an increased motivation of the patients to move around in the VE. The questions were answered by the participants on a visual analog scale. Further, the participants were asked to rank the 4 gaming conditions according to their preferences.

The physiotherapists' questionnaire assessed the acceptability of the VR HMD as a therapeutic tool for their patient. Questions included their ratings of the participants' movement behavior, level of support needed, and engagement during the VR session on a 5-point Likert scale. They were also asked about their opinion on using a VR HMD as a therapy tool and the advantages, disadvantages, and potential problems of using a VR HMD with their patients.

Furthermore, the absolute position of the VR HMD in the room was logged with a sampling rate of 50 Hz. Based on this data, the covered horizontal and vertical distance per gaming condition and participant was calculated.

Statistical Analysis

Participants' characteristics and covered horizontal and vertical distances are presented using descriptive statistics. Questionnaire responses are illustrated with frequencies, medians, and IQRs. To quantify potential differences between the 3 games, we tested the horizontal and vertical distances covered during the 3 gaming conditions for normal distribution and performed repeated measures ANOVA with Bonferroni-corrected post hoc tests. Calculated effect sizes were based on the *z* values of the nonparametric Wilcoxon rank sum tests. We interpreted $\omega^2>0.01$ as small, $\omega^2>0.06$ as moderate, and $\omega^2>0.14$ as large effect [28].

Results

Participant Characteristics

In total, 21 children and adolescents with a median age of 12.1 (IQR 5.5) years, of which two-thirds had a congenital neuromotor disorder, participated in this study (Table 1). The walking abilities of our study population were on a high level, as more than half (67%, n=14) of the participants were able to walk independently without an assistive device (FMS≥5), and 18 (86%) could walk outdoors at least for short distances (FAQ≥6). In all, 2 participants performed the VR HMD test in their wheelchair, 7 used a walking device, and 6 needed the help of their therapist, mainly in the form of supervision. No youth reported any signs of cybersickness in the VRSQ, neither after 5 nor 20 minutes of VR HMD exposure. Further patient characteristics are presented in Table 1.

Ratings on the usability of the VR HMD, user experience, and acceptability of the immersive VE are described in the following paragraphs.

XSL•FO
RenderX

Table 1. Descriptive characteristics of the study participants (N=21).

ID	Sex	Age (year)	Height (cm)	Diagnosis[a]	Glasses	FMS[b]	FAQ[c]	Mobility aid[d]
1	F[e]	8.2	120	Bilateral spastic-dystonic CP[f] (II)	Yes	6, 6, 5	9	None
2	M[g]	12.3	152	Stroke	No	6, 6, 6	10	None
3	F	11.5	147	Bilateral spastic CP (III) with lower limb surgery	No	1, 1, 1	4	Wheelchair
4	M	15.2	167	Bilateral spastic CP (II)	No	5, 5, 5	8	None
5	M	9.0	121	Meningomyelocele with lower limb surgery	No	5, 2, 2	8	Posterior walker
6	F	9.4	147	Cerebral encephalopathy	No	6, 6, 6	9	None
7	M	17.7	183	Stroke	No	5, 5, 5	7	None
8	M	11.6	146	Traumatic brain injury	No	6, 6, 6	10	None
9	M	7.3	129	Stroke	No	3, 3, 3	8	Crutches
10	F	14.6	169	Traumatic brain injury	Yes	5, 5, NA[h]	6	None
11	F	10.6	127	Unclear, superimposed disease with spastic-dystonic gait disorder	Yes	5, 5, 5	9	None
12	M	6.8	114	Meningomyelocele with lower limb surgery	Yes	2, 1, 1	4	Posterior walker
13	M	12.1	150	Status post septic shock with ischemic cerebral lesions	Yes	5, 3, 2	7	None
14	M	9.1	140	Bilateral spastic CP (II) with lower limb surgery	Yes	2, 2, 1	7	Posterior walker
15	M	14.6	160	Meningomyelocele with lower limb surgery	Yes	2, 2, 1	6	Posterior walker
16	M	14.4	167	Bilateral spastic CP (III) with lower limb surgery	No	2, 1, 1	3	Wheelchair
17	F	16.0	168	Friedreich ataxia with spondylodesis T4-L3	No	2, 2, NA	6	Anterior walker
18	F	12.4	162	Meningomyelocele	Yes	6, 6, 6	10	None
19	F	15.5	161	Bilateral spastic CP (II) with lower limb surgery	No	5, 3, 3	7	Crutches
20	F	7.9	122	Unilateral spastic CP (I) with lower limb surgery	Yes	6, 6, 6	10	None
21	F	13.0	172	Congenital ataxia	No	5, 5, 5	9	None

[a]In children and adolescents diagnosed with cerebral palsy, the Gross Motor Classification System Level is given in parentheses.

[b]FMS: Functional Mobility Scale at 5, 50, and 500 m.

[c]FAQ: Gillette Functional Assessment Questionnaire walking scale.

[d]Mobility aid used during virtual reality head-mounted display testing.

[e]F: female.

[f]CP: cerebral palsy.

[g]M: male.

[h]NA: not assessed.

Usability

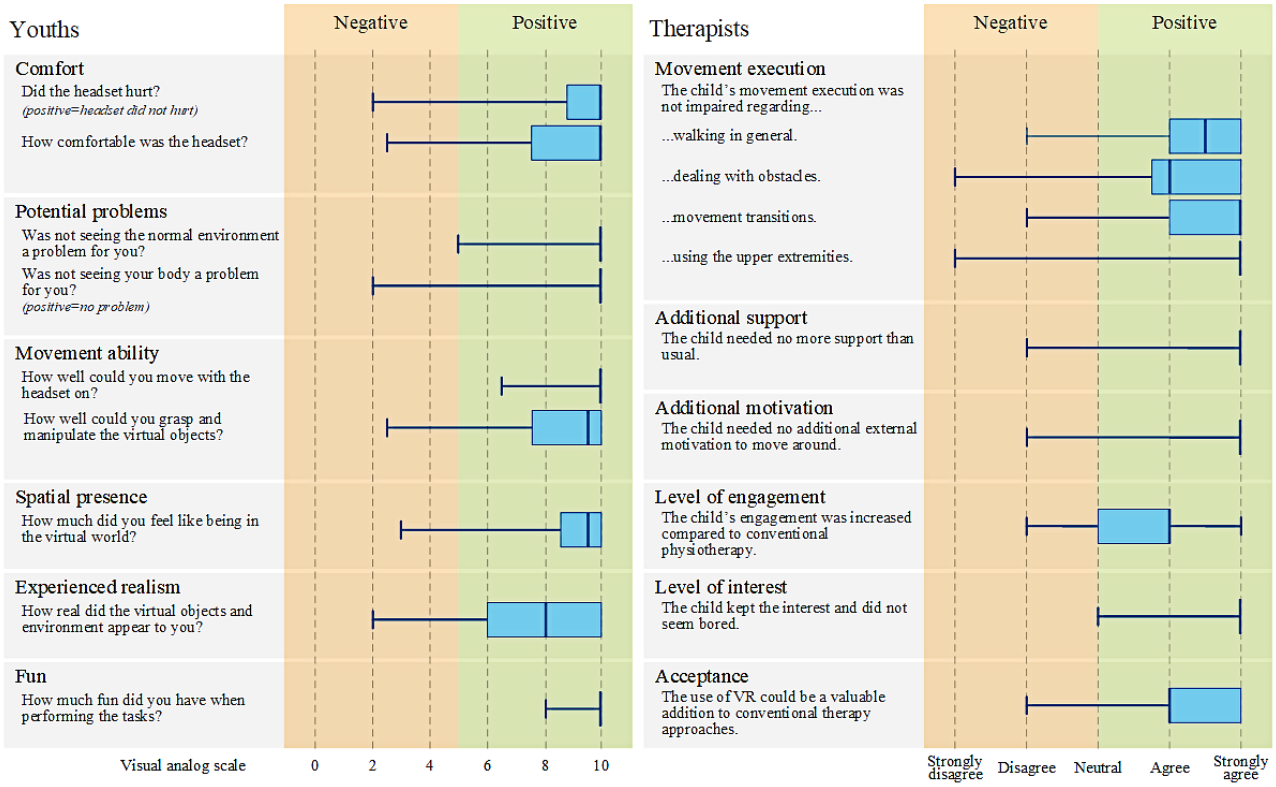
The participants rated the comfort of the VR HMD after 20 minutes of exposure time almost exclusively as positive. Only 3 patients reported that the VR HMD caused uncomfortable pressure on the back of the head (ID 14) or nose (IDs 16 and 17). Only 1 participant (ID 17) stated that not seeing her own body while moving around was a problem. None of the participants regarded it as unfavorable that the VR HMD blocked their view of the natural environment. Consequently, the youths did not feel hindered in their movement abilities by the VR HMD, except for 2 patients (IDs 5 and 13), who reported that grasping and manipulating the virtual objects was somewhat difficult for them.

User Experience

The majority (90%, 19/21) of the participants felt that the VE and objects they saw were real (ie, really present). They also rated their feeling of being present in the VE as very high (median 9.5, IQR 1.5; Figure 2). When asked how much fun they had performing the tasks in the VE, 19 (90%) of the 21 patients gave the maximum score of 10. The badger and the possibility to feed it was most (71%, 15/21) participants' favorite part in the VE.

XSL•FO
RenderX

Figure 2. Youths' and therapists' ratings of various parameters regarding the usability of the virtual reality head-mounted display, the user experience, and the acceptability of the immersive virtual environment. VR: virtual reality.



Acceptability

According to the physiotherapists, 5 patients needed more support than usual when moving around with the VR HMD, mainly in the form of supervision (IDs 7, 17, and 20) or assistance with their mobility aids (IDs 3 and 9). Otherwise, the VR HMD hardly affected the participants' movement behavior. The physiotherapists rated their patients' level of engagement during the VR session higher than in conventional physiotherapy (62%, 13/21; Figure 2).

They considered the immersive VE a valuable complement to conventional therapy methods for training everyday walking tasks in all but 3 participants (IDs 1, 12, and 21). Increased motivation, movement variations, repetitions, concentration level, playfulness, sense of achievement, joy of discovery, competition possibility with games, dual-task training, and reduced fear of movement were favorable factors mentioned by the physiotherapists. Potential problems or disadvantages identified by the therapists were the increased difficulty in handling the mobility aids when reaching for virtual objects, difficulty to work on gait quality, lack of haptic feedback, nonvisibility of the feet, weight of the VR HMD, and the VE not being challenging enough for some patients.

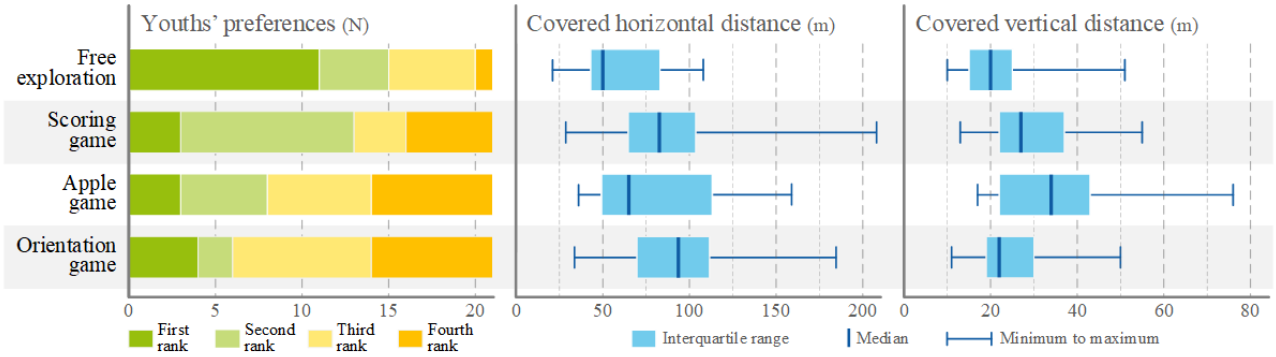
VE Conditions

The *free exploration mode* received the best rating of the 4 conditions, followed by the *scoring game* (Figure 3). The participants covered the most horizontal distance in the *scoring game* (median 94 m; range 29-208 m). However, the repeated measures ANOVA showed no statistically significant difference between the 3 games ($F_{2,34}$=2.60; P=.09).

The *apple collecting game* caused the largest covered vertical distance (median 34 m; range 17-76 m). The repeated measures ANOVA determined a statistically significant difference between the 3 games ($F_{2,34}$=19.31; P<.001; ω^2=0.5). Bonferroni-adjusted post hoc analyses revealed that the participants covered significantly more vertical distance in the *apple game* than the *scoring game* (P=.008) and *orientation game* (P<.001) as well as in the *scoring game* than the *orientation game* (P=.048).

XSL•FO
RenderX

Figure 3. Overview of the 4 application modes by participant preferences and distance covered in the horizontal and vertical plane.



Discussion

Principal Findings

We explored an immersive VE with different game conditions experienced through a VR HMD to train everyday life walking activities in a pediatric neurorehabilitation setting. The participants' and therapists' ratings regarding usability, user experience, and acceptability were very positive. The youths enjoyed moving around in the VE and experienced enormous fun—regardless of their age. Almost all participants felt comfortable with the VR HMD, and they did not notice any restrictions in their freedom of movement.

The participants reported a high sense of presence, indicating that they felt like being physically and spatially located in the VE. The sense of presence is a crucial feature of a VR application, as it directly influences users' enjoyment during a VR game [29]. This high sense of presence might have also substantially contributed to the fact that the youths were not concerned by their blocked view of the natural surroundings. Remarkably, only 1 participant (ID17, diagnosed with Friedreich ataxia) considered the inability to see her body as problematic. The patients perceived moving around with the VR HMD as similar or equal to that in everyday life, but they rated their experience of activities involving hand and finger movements as slightly lower. Furthermore, 2 participants even experienced these activities as challenging, despite the visual representation of their hands in the VE. In fact, the integrated hand tracking option of the Oculus Quest caused some difficulties, as the hands are only detected as long as they are in the VR HMD cameras' field of view. The partial concealment of the hand due to the patients' hair or mobility aid and too fast hand and finger movements while grasping were other causes for the interrupted hand tracking of the device. Despite these challenges, the hand tracking feature enables an intuitive use of the hands in the VE, and the participants remained largely unperturbed and tried patiently to grasp the virtual objects.

From the physiotherapists' point of view, the VR HMD hardly affected their patients' movements, neither with respect to walking or crossing obstacles nor regarding movement transitions, using the upper extremities, or dealing with their walking devices. This finding is essential since a negative influence of the VR HMD on the patients' movement execution would be unfavorable for efforts to validate its future clinical application. Furthermore, in contrast to our first study [21], therapists did not feel that their patients needed more support with the VR HMD than in standard therapy. This disparity in judgment might be because the therapists were already more familiar with this new technology and had more confidence that their patients could move safely even though they could not see the real environment.

Similar to the youths' feedback, the therapists' reported that their patients were more dedicated than during regular therapy sessions, for example, by being more interested and focused on the tasks or showing more perseverance. Further, they moved around without additional external motivation and did not seem to lose interest during the 20-minute test session. As motivation is an indispensable factor for active participation during the therapy, these findings are another indication of the promising option of immersive VR as a tool for movement therapy. Additionally, our results align with the findings of other studies that uniformly report the fun and enjoyment of various pediatric [17,18,20,21,30] and adult patient populations [15,16] when experiencing VR as a therapy instrument.

Despite the relatively high weight of the VR HMD, our patients could wear it for 20 minutes with almost no problems. Furthermore, we observed that the VR HMD is also suitable for different activities, such as walking, stooping, kneeling, crawling, or even running. This is new information, as in pediatric populations, the usability and acceptance of VR HMDs have so far only been tested in static situations [17-20] or when moving around for a short duration of a few minutes [21]. Studies with healthy adults and patients with stroke, multiple sclerosis, and Parkinson disease have demonstrated that VR HMDs could successfully be used while the participants were walking on a treadmill [15,16]. However, treadmill walking is much more restricted and controlled than moving around in an open space as our patients did. Everyday life activities usually happen in an unrestricted way and require a range of movement variations. Thus, our approach takes gait therapy even a step closer to daily life.

Whether a VR HMD is a helpful complement to conventional gait therapy methods depends on the patient's individual abilities, preferences, and the specific therapy goals. Although the missing haptic feedback of not successfully mastering obstacles or the nonvisibility of the feet can make training on gait quality more difficult for some children and adolescents, this absence could help others increase their concentration on

XSL•FO
RenderX

their proprioceptive input. Additionally, the possibility to train all activities on even ground might be helpful for some patients. However, the challenge could be too small for others because of an absent uneven terrain. According to the therapists, the patients often seemed to stay more concentrated for longer time periods and demonstrated more endurance and perseverance in the VE than in conventional therapy sessions. In the study of Lai et al [30], the participants reported that their exercise amount had been significantly facilitated by the immersive and enjoyable character of the VR HMD applications. The application's novelty, the numerous discovery opportunities in the VE, and the games' playfulness and competitive nature may, therefore, also have positively contributed to the positive findings in our study.

Limitations and Future Considerations

The mobility level of our participants was at a high level, with 15 being community walkers with an FAQ level of at least 7. The comments of the physiotherapists, who considered our VE not to be the right approach to address some of these patients' therapeutic goals, indicate that the VE was not challenging enough for patients who can master uneven grounds without any assistance (FAQ≥9), which applied to 8 participants of our study population. The introduction of various difficulty levels, dual-task training, and other VEs with more challenging tasks are possible future solutions to provide an adequate training level for these patients with higher functioning. As we did not compare the visual design of the current VE to others, we cannot comment on the effect of our particular visual design or VE choices.

Although we randomized the order of the 3 games, this was not the case for the *free exploration mode*. The participants always started with this condition as a type of familiarization with the VR HMD and VE. Therefore, we could not include the *free exploration mode* in the analysis regarding the covered distances. Neither can we rule out that the majority of the participants preferred the *free exploration mode*, because they experienced the VE with this condition in the first place. Nevertheless, it seems advisable to provide new VEs and interaction features always with an additional noncompetitive option, whether for familiarization or for children and adolescents who prefer noncompetitive games.

Our participant group was substantially heterogeneous in terms of age and motor abilities. Although it would have been interesting to analyze the impact of these characteristics, our sample size precluded forming any subgroups for further analyses. Additionally, as the study was cross-sectional, we cannot draw any conclusions about the impact of our VR HMD approach on the change in patients' functional walking abilities or how their motivation would develop over the long term. Further, the assessment of the participants' qualitative movement behavior was solely based on the subjective ratings of the youths themselves and their therapists.

Consequently, in a current study, we use 3D gait analysis to record spatiotemporal and kinematic parameters to compare the patients' movements when performing similar activities in real environments and VEs. Furthermore, we aim to develop and implement a foot tracking option in the VE. This option would provide patients with visual feedback on their feet's position and create further interaction possibilities, which would further help improve the VR experience. Last, the implementation of a movement therapy using a VR HMD is required in a clinical setting to evaluate its value and effect on patients' motivation and movement skills.

Conclusions

This study demonstrates the usability of an immersive VE delivered through a VR HMD in children and adolescents with neuromotor impairments performing everyday life walking activities. Furthermore, participants' and therapists' ratings regarding user experience and acceptability and the application's high motivational impact support its development as a future tool for movement therapy in pediatric neurorehabilitation. However, in the light of the current generation's rich gaming experience, choices, variation, difficulty levels, and other typical gaming features seem to be indispensable properties for successfully implementing the VR HMD as a therapy tool.

Acknowledgments

We are grateful to the children, adolescents, and therapists for the time and effort they dedicated to this project. Further, we thank Augment IT by Netcetera for the valuable collaboration and their efforts and support in the development of the virtual environment. This project was supported by the J&K Wonderland Foundation, Steinhausen, Switzerland; and the Promedica Foundation Chur, Switzerland. We also acknowledge the Children's Research Center of the University Children's Hospital Zurich.

Authors' Contributions

CAR contributed to conceptualization, methodology, investigation, formal analysis, and project coordination and wrote the original draft. AK contributed to conceptualization, methodology, investigation, formal analysis, reviewing, and editing. UK contributed to conceptualization, methodology, investigation, writing, reviewing, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video of pediatric neurorehabilitation patients experiencing and interacting with the immersive virtual environment.

XSL•FO
RenderX

[MP4 File (MP4 Video), 102150 KB - games_v10i3e38509_app1.mp4]

References

1. Moll LR, Cott CA. The paradox of normalization through rehabilitation: growing up and growing older with cerebral palsy. Disabil Rehabil 2013 Jul;35(15):1276-1283. [doi: 10.3109/09638288.2012.726689] [Medline: 23066918]
2. Maier M, Ballester BR, Verschure PFMJ. Principles of neurorehabilitation after stroke based on motor learning and brain plasticity mechanisms. Front Syst Neurosci 2019 Dec 17;13:74 [FREE Full text] [doi: 10.3389/fnsys.2019.00074] [Medline: 31920570]
3. Massetti T, da Silva TD, Crocetta TB, Guarnieri R, de Freitas BL, Bianchi Lopes P, et al. The clinical utility of virtual reality in neurorehabilitation: a systematic review. J Cent Nerv Syst Dis 2018;10:1179573518813541 [FREE Full text] [doi: 10.1177/1179573518813541] [Medline: 30515028]
4. Levac DE, Huber ME, Sternad D. Learning and transfer of complex motor skills in virtual reality: a perspective review. J Neuroeng Rehabil 2019 Oct 18;16(1):121 [FREE Full text] [doi: 10.1186/s12984-019-0587-8] [Medline: 31627755]
5. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. Front Psychol 2018;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
6. Cano Porras D, Siemonsma P, Inzelberg R, Zeilig G, Plotnik M. Advantages of virtual reality in the rehabilitation of balance and gait: systematic review. Neurology 2018 May 29;90(22):1017-1025. [doi: 10.1212/WNL.0000000000005603] [Medline: 29720544]
7. Shen J, Johnson S, Chen C, Xiang H. Virtual reality for pediatric traumatic brain injury rehabilitation: a systematic review. Am J Lifestyle Med 2020 Feb 06;14(1):6-15 [FREE Full text] [doi: 10.1177/1559827618756588] [Medline: 31903073]
8. Chen Y, Fanchiang HD, Howard A. Effectiveness of virtual reality in children with Cerebral Palsy: a systematic review and meta-analysis of randomized controlled trials. Phys Ther 2018 Jan 01;98(1):63-77 [FREE Full text] [doi: 10.1093/ptj/pzx107] [Medline: 29088476]
9. Levac D, Rivard L, Missiuna C. Defining the active ingredients of interactive computer play interventions for children with neuromotor impairments: a scoping review. Res Dev Disabil 2012;33(1):214-223. [doi: 10.1016/j.ridd.2011.09.007] [Medline: 22093667]
10. Ravi DK, Kumar N, Singhi P. Effectiveness of virtual reality rehabilitation for children and adolescents with cerebral palsy: an updated evidence-based systematic review. Physiotherapy 2017 Sep;103(3):245-258. [doi: 10.1016/j.physio.2016.08.004] [Medline: 28109566]
11. Lei C, Sunzi K, Dai F, Liu X, Wang Y, Zhang B, et al. Effects of virtual reality rehabilitation training on gait and balance in patients with Parkinson's disease: a systematic review. PLoS One 2019;14(11):e0224819 [FREE Full text] [doi: 10.1371/journal.pone.0224819] [Medline: 31697777]
12. de Rooij IJM, van de Port IGL, Meijer JWG. Effect of virtual reality training on balance and gait ability in patients with stroke: systematic review and meta-analysis. Phys Ther 2016 Dec;96(12):1905-1918. [doi: 10.2522/ptj.20160054] [Medline: 27174255]
13. Perez-Marcos D. Virtual reality experiences, embodiment, videogames and their dimensions in neurorehabilitation. J Neuroeng Rehabil 2018 Nov 26;15(1):113 [FREE Full text] [doi: 10.1186/s12984-018-0461-0] [Medline: 30477527]
14. Delgado F, Der Ananian C. The use of virtual reality through head-mounted display on balance and gait in older adults: a scoping review. Games Health J 2021 Feb;10(1):2-12. [doi: 10.1089/g4h.2019.0159] [Medline: 32598189]
15. Kim A, Darakjian N, Finley JM. Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. J Neuroeng Rehabil 2017 Feb 21;14(1):16 [FREE Full text] [doi: 10.1186/s12984-017-0225-2] [Medline: 28222783]
16. Winter C, Kern F, Gall D, Latoschik ME, Pauli P, Käthner I. Immersive virtual reality during gait rehabilitation increases walking speed and motivation: a usability evaluation with healthy participants and patients with multiple sclerosis and stroke. J Neuroeng Rehabil 2021 Apr 22;18(1):68 [FREE Full text] [doi: 10.1186/s12984-021-00848-w] [Medline: 33888148]
17. Phelan I, Furness PJ, Dunn HD, Carrion-Plaza A, Matsangidou M, Dimitri P, et al. Immersive virtual reality in children with upper limb injuries: findings from a feasibility study. J Pediatr Rehabil Med 2021;14(3):401-414 [FREE Full text] [doi: 10.3233/PRM-190635] [Medline: 34151871]
18. Tennant M, McGillivray J, Youssef GJ, McCarthy MC, Clark T. Feasibility, acceptability, and clinical implementation of an immersive virtual reality intervention to address psychological well-being in children and adolescents with cancer. J Pediatr Oncol Nurs 2020;37(4):265-277. [doi: 10.1177/1043454220917859] [Medline: 32536320]
19. Lai B, Powell M, Clement AG, Davis D, Swanson-Kimani E, Hayes L. Examining the feasibility of early mobilization with virtual reality gaming using head-mounted display and adaptive software with adolescents in the pediatric intensive care unit: case report. JMIR Rehabil Assist Technol 2021 May 27;8(2):e28210 [FREE Full text] [doi: 10.2196/28210] [Medline: 34042602]
20. Shen J, Xiang H, Luna J, Grishchenko A, Patterson J, Strouse RV, et al. Virtual reality–based executive function rehabilitation system for children with traumatic brain injury: design and usability study. JMIR Serious Games 2020 Aug 25;8(3):e16947 [FREE Full text] [doi: 10.2196/16947] [Medline: 32447275]

XSL•FO
RenderX

21. Ammann-Reiffer C, Keller U, Kläy A, Meier L, van Hedel HJA. Do youths with neuromotor disorder and their therapists prefer a mixed or virtual reality head-mounted display? J Rehabil Med Clin Commun 2021;4:1000072 [FREE Full text] [doi: 10.2340/20030711-1000072] [Medline: 34804416]
22. Ammann-Reiffer C, Graser JV. Walking activities beyond gait training: priorities in everyday life for parents and adolescents in pediatric neurorehabilitation. J Pediatr Rehabil Med 2022;15(2):311-321. [doi: 10.3233/PRM-201513] [Medline: 35527582]
23. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
24. Ammann-Reiffer C, Bastiaenen CHG, Klöti C, van Hedel HJA. Concurrent validity of two gait performance measures in children with neuromotor disorders. Phys Occup Ther Pediatr 2019;39(2):181-192. [doi: 10.1080/01942638.2017.1420003] [Medline: 29336642]
25. Knight JF, Baber C. A tool to assess the comfort of wearable computers. Hum Factors 2005;47(1):77-91. [doi: 10.1518/0018720053653875] [Medline: 15960088]
26. Schubert TW. The sense of presence in virtual environments: a three-component scale measuring spatial presence, involvement, and realness. Zeitschrift für Medienpsychologie 2003 Apr;15(2):69-71. [doi: 10.1026//1617-6383.15.2.69]
27. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence 1998 Jun;7(3):225-240 [FREE Full text] [doi: 10.1162/105474698565686]
28. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
29. Shafer DM, Carbonara CP, Korpi MF. Factors affecting enjoyment of virtual reality games: a comparison involving consumer-grade virtual reality technology. Games Health J 2019 Feb;8(1):15-23. [doi: 10.1089/g4h.2017.0190] [Medline: 30199273]
30. Lai B, Davis D, Narasaki-Jara M, Hopson B, Powell D, Gowey M, et al. Feasibility of a commercially available virtual reality system to achieve exercise guidelines in youth with spina bifida: mixed methods case study. JMIR Serious Games 2020 Sep 03;8(3):e20667 [FREE Full text] [doi: 10.2196/20667] [Medline: 32880577]

Abbreviations

FAQ: Gillette Functional Assessment Questionnaire walking scale
FMS: Functional Mobility Scale
HMD: head-mounted display
SCR: Swiss Children's Rehab
VE: virtual environment
VR: virtual reality
VRSQ: Virtual Reality Sickness Questionnaire

Edited by N Zary; submitted 05.04.22; peer-reviewed by B Lai, J Luna; comments to author 13.05.22; accepted 12.06.22; published 14.07.22.

Please cite as:
Ammann-Reiffer C, Kläy A, Keller U
Virtual Reality as a Therapy Tool for Walking Activities in Pediatric Neurorehabilitation: Usability and User Experience Evaluation
JMIR Serious Games 2022;10(3):e38509
URL: https://games.jmir.org/2022/3/e38509
doi:10.2196/38509
PMID:35834316

©Corinne Ammann-Reiffer, Andrina Kläy, Urs Keller. Originally published in JMIR Serious Games (https://games.jmir.org), 14.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Design, Development, and Testing of a Virtual Reality Device for Upper Limb Training in People With Multiple Sclerosis: Single-Center Feasibility Study

Alon Kalron[1,2], PT, PhD; Lior Frid[2], BEd; Iliya Fonkatz[2], MA; Shay Menascu[2,3], MD; Mark Dolev[2], MD; David Magalashvili[2], MD; Anat Achiron[2,3], MD, PhD

[1]Department of Physical Therapy, School of Health Professions, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

[2]Multiple Sclerosis Center, Sheba Medical Center, Tel Hashomer, Israel

[3]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

Corresponding Author:
Alon Kalron, PT, PhD
Department of Physical Therapy
School of Health Professions, Sackler Faculty of Medicine
Tel-Aviv University
Ramat-Aviv
Tel-Aviv, 39040
Israel
Phone: 972 036405430
Email: alonkalr@post.tau.ac.il

Abstract

Background: Multiple sclerosis (MS) is a common nontraumatic, neurological, disabling disease that often presents with upper limb dysfunction. Exercise training has resulted in improvement for patients; however, there can be a lack of compliance due to access because of location and lack of MS experts. Virtual reality (VR) is a promising technology that can offer exercise therapy/rehabilitation at a distance. This type of remote training can be motivational and effective for patients with MS and can improve range of motion and muscle strength for those with upper limb dysfunction.

Objective: The aim of this study is to evaluate the safety and feasibility of the XRHealth software and the Oculus Rift Station for patients with MS with upper limb motor dysfunction.

Methods: A single-center, prospective, feasibility study was conducted with patients with MS who had upper limb motor dysfunction. Patients participated in a single 45-minute digital environment session with VR and completed a questionnaire about the quality of the training and fatigability. The clinician also completed a questionnaire to evaluate the suitability and safety of the training.

Results: Overall, 30 patients were enrolled between the ages of 20 and 81 years. Patients reported that the training sessions within the digital environment were helpful, challenging, fun, and simple to understand, and that they would be willing to repeat the sessions again. The physical therapist that oversaw the patients reported that the training was suitable for 87% (n=26) of the patients. Anticipated adverse events were fatigue, temporary dizziness, and temporary nausea. The operator complications included that the cable of the head-mounted display interrupted the training (n=2, 7%) and fatigue that caused cessation of the VR training session (n=2, 7%). No serious adverse events were reported.

Conclusions: These preliminary results demonstrated that the use of the XRHealth software and Oculus Rift Station platform is feasible, safe, and engaging for patients, and has the potential to improve the functionality of the upper limbs in patients with MS. This study provides support for future studies of implementing a series of training sessions with virtual reality in a home-based environment.

(JMIR Serious Games 2022;10(3):e36288) doi:10.2196/36288

KEYWORDS

virtual reality; rehabilitation; feasibility; upper limb; multiple sclerosis

XSL•FO
RenderX

Introduction

Background

Multiple sclerosis (MS) is a chronic progressing demyelinating disease of the central nervous system [1]. The age at which MS is generally diagnosed is between the ages of 20-40 years, with a higher prevalence in female patients than male patients [2]. MS is the most common nontraumatic neurological disabling disease affecting young adults with an incidence of approximately 1 in 1000 [2], which affects approximately 2.8 million people worldwide (35.9 per 100,000 population) [3].

In patients with MS, upper limb dysfunction is common and occurs in approximately 50% to 76% of patients [4,5]. The presentation of upper limb dysfunction can be variable in patients with a range of physical symptoms presenting as tremor, lack of movement coordination, or muscle weakness [1,6,7]. These symptoms can lead to a decreased quality of life (QoL) due to poor health status, psychological difficulties, depressed mood and cognitive difficulties [8], increased mortality, restriction of activities [9,10], and dependency on others [11]. Some of the most frequent challenges that patients with MS face is dexterity, activities of daily living, and limitations in the upper limbs [12]. Treatment approaches for MS typically involve pharmacological intervention and rehabilitation treatments [13-15].

Symptoms of MS can range from visual impairment; dysarthria and dysphagia; impairment of coordination and balance; ataxia; sensory impairment; and bladder, bowel and sexual dysfunction [1,16-19]. Exercise training for MS and those with upper limb dysfunction has demonstrated significant benefit [20-24] and has been able to mitigate symptoms and decrease dependency on others [15,22,23].

The major reasons for lack of compliance in patients with MS are the inability to access exercise programs because patients live in a rural area or location that lack MS experts that support traditional upper limb exercise programs. Patient transportation and access can be limited due to time commitments, vehicle access, and simply not having the health and physical ability to travel to a hospital for physical rehabilitation. Another major obstacle is the cost of care that is prohibitive for patients. The COVID-19 pandemic has highlighted the need to provide therapy/rehabilitation from a distance, mainly due to travel restrictions, access, and risk of infection [25]. Since rehabilitation can be especially difficult for the patient due to time and physical demands, loss of motivation and compliance often occur; therefore, it is critical to develop innovative new methods to keep patients motivated and improve QoL.

A method to circumvent these challenges and support exercise training in the home of patients and increase physical activity and functionality, thereby improving QoL, is telerehabilitation [26]. Therapy is delivered remotely with telerehabilitation outside traditional hospital settings, providing access to patients who have problems with mobility and travel restrictions. A promising new component is virtual reality (VR) for patients. VR provides a 3D-simulated environment with a computer interacting with a number of electronic accessories that can motivate patients to do repetitive tasks with sensory and visual feedback [27]. The level of immersion with VR may vary, ranging from nonimmersive (eg, video games) to fully immersive, where the user must have proper VR glasses or a head-mounted display such as the HTC Vive or Oculus Rift. There are different ideas on the classification of VR to date, and a classification remains to be standardized between studies [28]. Immersive VR has been applied to psychological therapy for many disorders including posttraumatic stress disorder, borderline personality disorder, schizophrenia, and other psychological disorders [29]. Patients can practice functional tasks such as shopping or crossing the street in a virtual environment [27]. The benefits of VR have been established in multiple applications including pain management and reduction [30-34], mental health support [33-36], and rehabilitation [32]. In MS, the application of VR has improved balance, posture, motor control, coordination, and gait [37-40] as well as improved motor function in the upper limbs [41]. A recent study demonstrated an improvement in manual dexterity for patients with MS using game-based VR video capture training [42]. The controlled environment of immersive VR can lead to improved patient outcomes compared with conventional therapy in different types of motor impairments including hemiparesis caused by Parkinson disease, MS, cerebral palsy, and stroke [28,29,41,43]. Despite the promising results, VR is not widely implemented, resulting in an unmet medical need for a growing patient population.

Currently, XRHealth software is cleared for marketing in the United States by the Food and Drug Administration, Europe (EC certification), and Israel (AMAR). The software along with the Oculus Rift Station guides patients through exercises according to an individualized treatment plan from their medical practitioner. The Oculus Rift station has touch controllers and provides *df* for the movement inside a virtual environment, which includes high-end tracking of both the head and arm movements with 0.01 degrees in 90 Hz accuracy. Through the combination of the XRHealth software, the Oculus Rift Station, and a clinical therapist, the motion and movement kinetics of a patient can be tracked and provided to a clinician.

Objectives

The objective of this pilot study is to evaluate the combination of the XRHealth software with the Oculus Rift Station for safety and feasibility in patients with MS with upper limb motor dysfunction. The quality of the resulting data from this platform was assessed for optimization and serves as preliminary data for the development and design of future studies for VR training in people with MS.

Methods

Ethics Approval

This is a single-center safety and feasibility study based on a single 45-minute training session. All patients provided informed consent, and institutional review board approval was obtained at the Sheba Medical Center (SMC-5207-18). This study was conducted in the Sheba Multiple Sclerosis Center, Tel-Hashomer, Israel, between 2019 and 2020.

XSL•FO
RenderX

Patients

Patients were enrolled if they were 18 years or older, had moderate weakness in upper extremity muscles as defined by the British Medical Research Council including those with grade 4 in 2 muscle groups, or British Medical Research Council grade 3 in 1 muscle group. Patients were excluded if they had a hearing impairment, any orthopedic impairment restricting upper limb movements, or any cardiovascular disease preventing aerobic exercise.

VR Training Platform and Device

The intervention involved 2 activities performed once within the VR environment that would require both arm and shoulder movements either while seated (Figure S1A in Multimedia Appendix 1) or standing (Figure S1B in Multimedia Appendix 1) on the setup shown in Figure S1C in Multimedia Appendix 1. This was performed with the Oculus Rift and XRHealth software. The Oculus Rift headset is similar to goggles and, when worn, allows the patient to look in any direction naturally as if in the real world. In the Oculus Rift headset, there is a tracker that constantly analyzes head movement, allowing a clinician to make adjustments for the patient in real time [44]. The XRHealth software platform has three main components: a central console, a data portal, and a control panel. In the central console, a range of different motor, cognitive, and mental experiences in the form of games and exercises can be selected and serves as a "home environment" for all XRHealth applications. The data portal is a Health Insurance Portability and Accountability Act–compliant web-based platform that clinicians can use to track and analyze data from patients, and offers longitudinal data about patient performance and progress. The VR experience can be viewed and adjusted remotely and in real time by the clinician, and is supported by the control panel.

All assessments were overseen by a physical therapist with 2 years of experience in neurohabilitation. The first activity was an active shoulder range of motion (ROM) application called "Balloon Blast." In this VR environment, the patient uses both hands and is prompted to pop balloons with a swipe of a sword (Figure S2A in Multimedia Appendix 1). The area of shoulder activity is determined with an active ROM measurement (Figure S2B and C in Multimedia Appendix 1). After a 5-minute rest period, the patient performed the second activity, which is a motor cognitive training environment called "Color Match." Patients wear different colored virtual gloves on both hands and are then prompted to hit light bulbs with the hand that matches the bulb color (Figure S2D in Multimedia Appendix 1). The initial settings of the "Color Match" VR environment were set to a medium level for all patients and adjusted according to the patient's capabilities. Medium level was defined as 3.6 seconds between the moment the light turns on and then turns off for the patient to act.

Outcome Measures for Safety and Feasibility

Each patient answered 4 questionnaires for the assessment (Supplemental 1 in Multimedia Appendix 1), and the physical therapist who oversaw the session also completed a questionnaire (Supplemental 2 in Multimedia Appendix 1). The lowest score was 1 and the highest was 4. Patient answers were averaged together for a final result. Patient questionnaires aimed to evaluate patient experience in relation to the VR platform for optimization and to inform the design of future studies. Questions included patient assessment on the level of difficulty, physical effects such as fatigue, engagement, and suitability of the VR training session.

Results

Patient Characteristics

A total of 30 patients with MS with upper limb motor dysfunction were enrolled and completed the training. Of those, 50% (n=15) were female patients, the mean age was 50.8 (range 20.3-81.2) years, and the mean Expanded Disability Status Scale score was 5.4 (Table 1).

Table 1. Patient demographic and clinical characteristics.

Characteristic	Patients (N=30), n (%)	Value	
		Mean (SD)	Range
Age (years)	N/A[a]	50.74 (13.6)	20.3-81.2
Female, n (%)	15 (50)	N/A	N/A
Multiple sclerosis type, n (%)		N/A	N/A
Primary progressive multiple sclerosis	3 (10)		
Clinically isolated syndrome	3 (10)		
Relapsing and remitting multiple sclerosis	14 (47)		
Secondary progressive multiple sclerosis	10 (33)		
Disease duration (years)	N/A	18.6 (12.8)	0.6-58
Expanded Disability Status Scale	N/A	5.4 (1.7)	1.5-8.0

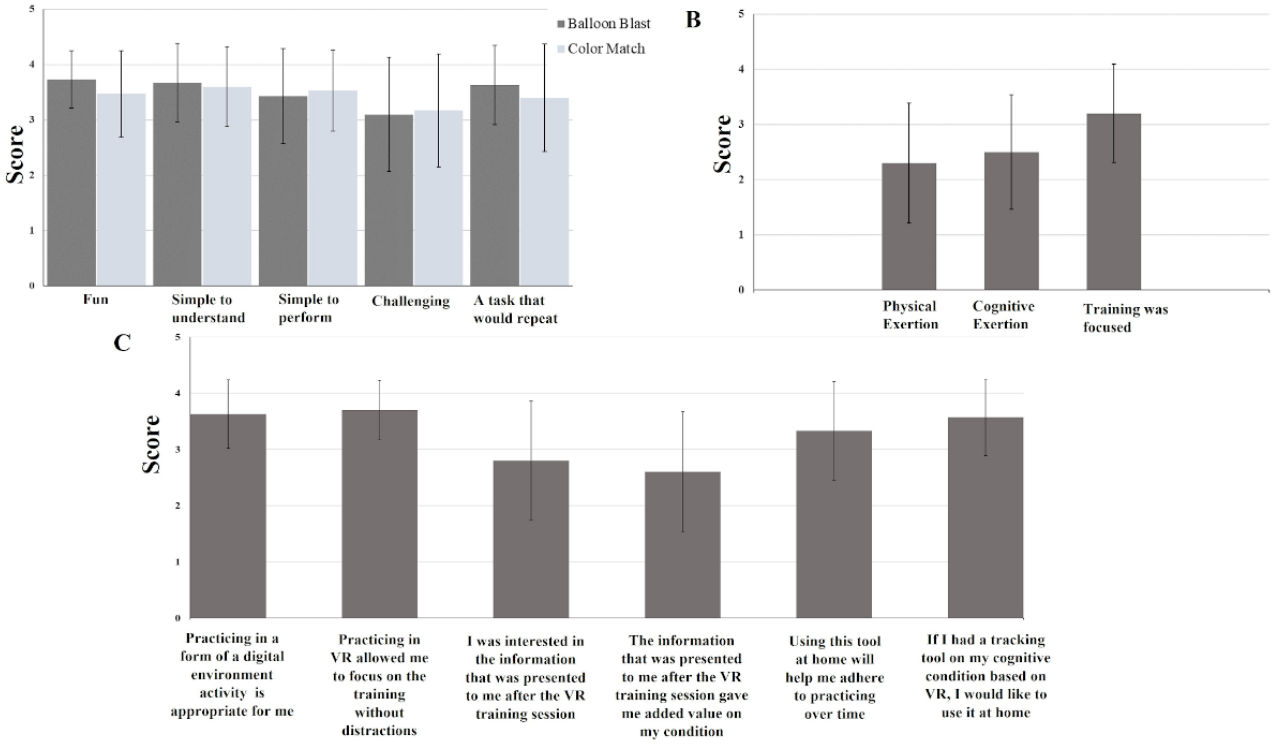
[a]N/A: not applicable.

XSL•FO
RenderX

Patient Satisfaction

Patients reported in the questionnaires that the "Balloon Blast" activity was fun (mean 3.73, SD 0.52), simple to understand (mean 3.67, SD 0.71), simple to perform (mean 3.43, SD 0.86), challenging (mean 3.10, SD 1.03), and a task that they would repeat (mean 3.63, SD 0.72; Figure 1A). In the "Color Match" activity, patients reported that the activity was fun (mean 3.47, SD 0.78), simple to understand (mean 3.60, SD 0.72), simple to perform (mean 3.53, SD 0.73), challenging (mean 3.17, SD 1.02), and a task that they would repeat (mean 3.40, SD 0.97; Figure 1A).

Figure 1. Questionnaire results: (A) self-reported satisfaction questionnaire average, (B) the overall VR experience reported by patients, and (C) statements about the session and levels of patient agreeance. VR: virtual reality.



For overall VR experience satisfaction, patients reported that during the VR training session they felt physical exertion (mean 2.3, SD 1.09) and cognitive exertion (mean 2.5, SD 1.04), and that the training was focused (mean 3.2, SD 0.89; Figure 1B). Using the same scoring system, patients reported high levels of agreement with the following statements: practicing in a form of a digital environment activity is appropriate for me (mean 3.63, SD 0.61); practicing in VR allowed me to focus on the training without distractions (mean 3.70, SD 0.53); I was interested in the information that was presented to me after the VR training session (mean 2.8, SD 1.06); the information that was presented to me after the VR training session gave me added value on my condition (mean 2.6, SD 1.07); using this tool at home will help me adhere to practicing over time (mean 3.33, SD 0.88); and if I had a tracking tool on my cognitive condition based on VR, I would like to use it at home (mean 3.57, SD 0.68; Figure 1C). After completing the VR training sessions with the patients, the physical therapist reported that this was a suitable solution for people with MS (n=26, 87%).

Safety

Reported intraoperative complications included that the cable of the head-mounted display interrupted the training (n=2, 7%) and fatigue that caused cessation of the VR training session (n=2, 7%). There were no other intraoperative complications reported. Anticipated adverse events (AEs) according to a 4-point Likert scale were fatigue (mean 1.9, SD 0.96), temporary dizziness (mean 1.57, SD 0.77), and temporary nausea (mean 1.17, SD 0.46). Of the patients that reported fatigue, 6 patients experienced light fatigue, 9 patients reported moderate fatigue, and 1 patient reported severe fatigue resulting in a score of 1. Of those that reported dizziness, 7 patients reported light dizziness and 4 patients reported moderate dizziness. In patients that reported nausea, 3 reported light nausea and 1 reported moderate nausea. No other unanticipated AEs were reported. There were no serious AEs. All the AEs reported were resolved within a short period of time.

Discussion

Principal Results

Overall, patients found the VR environment training session helpful and would be willing to do more. The physical therapist determined that the session was suitable for people with MS. Physical exercise therapy for patients with MS is critical in maintaining movement, mental health, and the ability to be self-sufficient [21]. This study supports moving the platform into a home-based environment, which reduces the financial and travel burden on patients while also improving health and QoL [45-48]. Furthermore, the VR digital environment is a simple patient interface in which patients reported the programs to be engaging, challenging, simple, and fun. Patients had minimal risk, and the activities were safe and tolerable. There

XSL•FO
RenderX

were no serious AEs, and those that were reported were temporary.

The XRHealth platform is performed under medical supervision during training and allows the physical therapist to modify several parameters during a session that meets the specific needs of the patient. The types of assessments that can be performed include ROM, action time, response time, and omission and commission mistakes collected during a single 45-minute training session, which is then output as a graph by the XR software (Figure S2E in Multimedia Appendix 1). The ability to involve a clinician and obtain a number of data points to determine outcomes while patients are receiving care in a digital environment are substantial benefits.

Limitations

This study is not without limitations. First, the study did not include standard measures of upper limb function or manual dexterity. In addition, future adaptations could use the standardized System Usability Scale rather than the customized questionnaire used in this study based on similar publications in the field. Furthermore, this was a small study with a single VR training session. Future larger studies are needed to investigate whether scores derived from the VR system correlate with clinical upper limb measures in people with MS and other populations with central neurological damage such as stroke or Parkinson disease. A larger and longer study with more sessions would provide the efficacy data needed to support the use of a VR environment for physical therapy in people with MS.

Comparison With Prior Work

The use of VR technology has revealed improvements in the clinic for patients with upper limb dysfunction and stroke [49,50], and has also demonstrated benefit and favorable results in improving gait impairments and balance in people with MS [51]. Early research in 2006 examining the efficacy of VR showed improvements in gait control for patients with MS [52]. A small study of 5 patients with MS showed that VR plus passive robotic support yielded improved upper limb function and was well tolerated [53]. In 2015, another study of 30 people with MS demonstrated that VR had significant improvement in overall stability after 24 sessions [38]. There are few prior studies, and it is unclear if the use of VR training in people with MS can improve the QoL status, although these studies show a positive trend in assisting people with MS. Currently, the use of VR is limited in the clinic for patients with MS despite preliminary successes. The XRHealth platform therefore has the ability to bring this technology to a broader patient population.

At this time, there are small ongoing randomized controlled trials (RCTs) integrating VR into MS therapy for upper limb dysfunction. These include the Telerehabilitation and Multiple Sclerosis (TEAMS) pilot RCT (ClinicalTrials.gov NCT04032431) comparing conventional therapy to a home-based telerehabilitation VR program [54], a single-blinded RCT using serious games for upper limb rehabilitation (ClinicalTrials.gov NCT04171908) [42], and a randomized interventional study adding game-based VR exercises to conventional physiotherapy (ClinicalTrials.gov NCT04212689) [55]. This pilot study contributes to these studies in support of integrating VR for rehabilitation of people with MS with upper limb dysfunction.

Conclusions

In conclusion, the XRHealth medical solution of a VR digital environment and Oculus Rift platform is a feasible and safe training system for upper limb training in people with MS. These findings pave the way for future RCTs to examine the benefits of the XRHealth VR training compared with standard care to improve functionality of the upper limbs in people with MS. Additionally, it would be noteworthy to compare VR training at a clinical facility compared to VR training telehealth, both monitored by a physical therapist. Finally, future research should examine the psychometric values of the outcome measures produced by the XRHealth platform in patients with upper limb dysfunctions.

Acknowledgments

JetPub Scientific Communications, LLC was supported by funding from XRHealth, and provided drafts and editorial assistance to the authors during preparation of this manuscript.

Authors' Contributions

AK and AA conceptualized the study. AK validated the study. AK and AA created the visualizations for the study. AK reviewed and edited the manuscript. LF was the project administrator. LF and IF preformed the investigation. SM, MD, and DM provided the resources for the study. AA supervised the study.

Conflicts of Interest

AA is a clinical advisor for XRHealth. All other authors have nothing to declare.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 2491 KB - games_v10i3e36288_app1.docx]

References

XSL•FO
RenderX

1. Compston A, Coles A. Multiple sclerosis. Lancet 2008 Oct 25;372(9648):1502-1517. [doi: 10.1016/S0140-6736(08)61620-7] [Medline: 18970977]
2. Koch-Henriksen N, Sørensen PS. The changing demographic pattern of multiple sclerosis epidemiology. Lancet Neurol 2010 May;9(5):520-532. [doi: 10.1016/S1474-4422(10)70064-8] [Medline: 20398859]
3. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. Mult Scler 2020 Dec;26(14):1816-1821 [FREE Full text] [doi: 10.1177/1352458520970841] [Medline: 33174475]
4. Holper L, Coenen M, Weise A, Stucki G, Cieza A, Kesselring J. Characterization of functioning in multiple sclerosis using the ICF. J Neurol 2010 Jan;257(1):103-113. [doi: 10.1007/s00415-009-5282-4] [Medline: 19756827]
5. Lamers I, Maris A, Severijns D, Dielkens W, Geurts S, Van Wijmeersch B, et al. Upper limb rehabilitation in people with multiple sclerosis: a systematic review. Neurorehabil Neural Repair 2016 Sep;30(8):773-793. [doi: 10.1177/1545968315624785] [Medline: 26747125]
6. Kister I, Bacon TE, Chamot E, Salter AR, Cutter GR, Kalina JT, et al. Natural history of multiple sclerosis symptoms. Int J MS Care 2013;15(3):146-158 [FREE Full text] [doi: 10.7224/1537-2073.2012-053] [Medline: 24453777]
7. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014 Jul 15;83(3):278-286 [FREE Full text] [doi: 10.1212/WNL.0000000000000560] [Medline: 24871874]
8. Johansson S, Ytterberg C, Claesson I, Lindberg J, Hillert J, Andersson M, et al. High concurrent presence of disability in multiple sclerosis. Associations with perceived health. J Neurol 2007 Jun;254(6):767-773. [doi: 10.1007/s00415-006-0431-5] [Medline: 17401746]
9. Bertoni R, Lamers I, Chen C, Feys P, Cattaneo D. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with multiple sclerosis. Mult Scler 2015 Oct;21(12):1566-1574. [doi: 10.1177/1352458514567553] [Medline: 25662346]
10. Kierkegaard M, Einarsson U, Gottberg K, von Koch L, Holmqvist L. The relationship between walking, manual dexterity, cognition and activity/participation in persons with multiple sclerosis. Mult Scler 2012 May;18(5):639-646. [doi: 10.1177/1352458511426736] [Medline: 21982871]
11. Yozbatiran N, Baskurt F, Baskurt Z, Ozakbas S, Idiman E. Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. J Neurol Sci 2006 Jul 15;246(1-2):117-122. [doi: 10.1016/j.jns.2006.02.018] [Medline: 16678208]
12. Kamm C, Heldner M, Vanbellingen T, Mattle H, Müri R, Bohlhalter S. Limb apraxia in multiple sclerosis: prevalence and impact on manual dexterity and activities of daily living. Arch Phys Med Rehabil 2012 Jun;93(6):1081-1085. [doi: 10.1016/j.apmr.2012.01.008] [Medline: 22464095]
13. Goldschmidt C, McGinley MP. Advances in the treatment of multiple sclerosis. Neurol Clin 2021 Feb;39(1):21-33 [FREE Full text] [doi: 10.1016/j.ncl.2020.09.002] [Medline: 33223085]
14. Sutliff MH, Bennett SE, Bobryk P, Halper J, Saslow LA, Skutnik LT, et al. Rehabilitation in multiple sclerosis: commentary on the recent AAN systematic review. Neurol Clin Pract 2016 Dec;6(6):475-479 [FREE Full text] [doi: 10.1212/CPJ.0000000000000318] [Medline: 28058205]
15. Döring A, Pfueller CF, Paul F, Dörr J. Exercise in multiple sclerosis -- an integral component of disease management. EPMA J 2011 Dec 24;3(1):2 [FREE Full text] [doi: 10.1007/s13167-011-0136-4] [Medline: 22738091]
16. DasGupta R, Fowler C. Bladder, bowel and sexual dysfunction in multiple sclerosis: management strategies. Drugs 2003;63(2):153-166. [doi: 10.2165/00003495-200363020-00003] [Medline: 12515563]
17. Henze T, Rieckmann P, Toyka KV, Multiple Sclerosis Therapy Consensus Group of the German Multiple Sclerosis Society. Symptomatic treatment of multiple sclerosis. Multiple Sclerosis Therapy Consensus Group (MSTCG) of the German Multiple Sclerosis Society. Eur Neurol 2006;56(2):78-105 [FREE Full text] [doi: 10.1159/000095699] [Medline: 16966832]
18. Cattaneo D, Jonsdottir J, Zocchi M, Regola A. Effects of balance exercises on people with multiple sclerosis: a pilot study. Clin Rehabil 2007 Sep;21(9):771-781. [doi: 10.1177/0269215507077602] [Medline: 17875557]
19. Darley F, Brown J, Goldstein N. Dysarthria in multiple sclerosis. J Speech Hear Res 1972 Jun;15(2):229-245. [doi: 10.1044/jshr.1502.229] [Medline: 5047862]
20. Pellegrino L, Coscia M, Muller M, Solaro C, Casadio M. Evaluating upper limb impairments in multiple sclerosis by exposure to different mechanical environments. Sci Rep 2018 Feb 01;8(1):2110. [doi: 10.1038/s41598-018-20343-y] [Medline: 29391520]
21. Halabchi F, Alizadeh Z, Sahraian MA, Abolhasani M. Exercise prescription for patients with multiple sclerosis; potential benefits and practical recommendations. BMC Neurol 2017 Sep 16;17(1):185 [FREE Full text] [doi: 10.1186/s12883-017-0960-9] [Medline: 28915856]
22. Motl R, Sandroff B. Benefits of exercise training in multiple sclerosis. Curr Neurol Neurosci Rep 2015 Sep;15(9):62. [doi: 10.1007/s11910-015-0585-6] [Medline: 26223831]
23. Swank C, Thompson M, Medley A. Aerobic exercise in people with multiple sclerosis: its feasibility and secondary benefits. Int J MS Care 2013;15(3):138-145 [FREE Full text] [doi: 10.7224/1537-2073.2012-037] [Medline: 24453776]

XSL•FO
RenderX

24. Pilutti LA, Platta ME, Motl RW, Latimer-Cheung AE. The safety of exercise training in multiple sclerosis: a systematic review. J Neurol Sci 2014 Aug 15;343(1-2):3-7. [doi: 10.1016/j.jns.2014.05.016] [Medline: 24880538]
25. Zasadzka E, Trzmiel T, Pieczyńska A, Hojan K. Modern technologies in the rehabilitation of patients with multiple sclerosis and their potential application in times of COVID-19. Medicina (Kaunas) 2021 May 30;57(6):549 [FREE Full text] [doi: 10.3390/medicina57060549] [Medline: 34070705]
26. Amatya B, Galea M, Kesselring J, Khan F. Effectiveness of telerehabilitation interventions in persons with multiple sclerosis: a systematic review. Mult Scler Relat Disord 2015 Jul;4(4):358-369. [doi: 10.1016/j.msard.2015.06.011] [Medline: 26195057]
27. Perez-Marcos D, Solazzi M, Steptoe W, Oyekoya O, Frisoli A, Weyrich T, et al. A fully immersive set-up for remote interaction and neurorehabilitation based on virtual body ownership. Front Neurol 2012;3:110. [doi: 10.3389/fneur.2012.00110] [Medline: 22787454]
28. Huygelier H, Mattheus E, Abeele VV, van Ee R, Gillebert C. The use of the term virtual reality in post-stroke rehabilitation: a scoping review and commentary. Psychol Belg 2021 Jun 03;61(1):145-162 [FREE Full text] [doi: 10.5334/pb.1033] [Medline: 34131490]
29. Elor A, Kurniawan S. The ultimate display for physical rehabilitation: a bridging review on immersive virtual reality. Frontiers Virtual Reality 2020 Nov 12;1:A. [doi: 10.3389/frvir.2020.585993]
30. Keefe FJ, Huling DA, Coggins MJ, Keefe DF, Rosenthal ZM, Herr NR, et al. Virtual reality for persistent pain: a new direction for behavioral pain management. Pain 2012 Nov;153(11):2163-2166 [FREE Full text] [doi: 10.1016/j.pain.2012.05.030] [Medline: 22770840]
31. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. Pain Manag Nurs 2015 Oct;16(5):685-691. [doi: 10.1016/j.pmn.2015.03.001] [Medline: 25972074]
32. Reid D. Virtual reality and the person-environment experience. Cyberpsychol Behav 2002 Dec;5(6):559-564. [doi: 10.1089/109493102321018204] [Medline: 12556119]
33. Park M, Kim D, Lee U, Na E, Jeon H. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019;10:505. [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
34. Diemer J, Alpers G, Peperkorn H, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 2015;6:26. [doi: 10.3389/fpsyg.2015.00026] [Medline: 25688218]
35. Srivastava K, Das RC, Chaudhury S. Virtual reality applications in mental health: challenges and perspectives. Ind Psychiatry J 2014;23(2):83-85 [FREE Full text] [doi: 10.4103/0972-6748.151666] [Medline: 25788795]
36. Rothbaum BO, Hodges L, Alarcon R, Ready D, Shahar F, Graap K, et al. Virtual reality exposure therapy for PTSD Vietnam Veterans: a case study. J Trauma Stress 1999 Apr;12(2):263-271. [doi: 10.1023/A:1024772308758] [Medline: 10378165]
37. Gutiérrez RO, Galán Del Río F, Cano de la Cuerda R, Alguacil Diego IM, González RA, Page JCM. A telerehabilitation program by virtual reality-video games improves balance and postural control in multiple sclerosis patients. NeuroRehabilitation 2013;33(4):545-554. [doi: 10.3233/NRE-130995] [Medline: 24029009]
38. Eftekharsadat B, Babaei-Ghazani A, Mohammadzadeh M, Talebi M, Eslamian F, Azari E. Effect of virtual reality-based balance training in multiple sclerosis. Neurol Res 2015 Jun;37(6):539-544. [doi: 10.1179/1743132815Y.0000000013] [Medline: 25668387]
39. Calabrò RS, Russo M, Naro A, De Luca R, Leo A, Tomasello P, et al. Robotic gait training in multiple sclerosis rehabilitation: can virtual reality make the difference? Findings from a randomized controlled trial. J Neurol Sci 2017 Jun 15;377:25-30. [doi: 10.1016/j.jns.2017.03.047] [Medline: 28477702]
40. Kalron A, Fonkatz I, Frid L, Baransi H, Achiron A. The effect of balance training on postural control in people with multiple sclerosis using the CAREN virtual reality system: a pilot randomized controlled trial. J Neuroeng Rehabil 2016 Mar 01;13:13. [doi: 10.1186/s12984-016-0124-y] [Medline: 26925955]
41. Webster A, Poyade M, Rooney S, Paul L. Upper limb rehabilitation interventions using virtual reality for people with multiple sclerosis: a systematic review. Mult Scler Relat Disord 2021 Jan;47:102610. [doi: 10.1016/j.msard.2020.102610] [Medline: 33212400]
42. Waliño-Paniagua CN, Gómez-Calero C, Jiménez-Trujillo MI, Aguirre-Tejedor L, Bermejo-Franco A, Ortiz-Gutiérrez RM, et al. Effects of a game-based virtual reality video capture training program plus occupational therapy on manual dexterity in patients with multiple sclerosis: a randomized controlled trial. J Healthc Eng 2019;2019:9780587. [doi: 10.1155/2019/9780587] [Medline: 31178989]
43. Iruthayarajah J, McIntyre A, Cotoi A, Macaluso S, Teasell R. The use of virtual reality for balance among individuals with chronic stroke: a systematic review and meta-analysis. Top Stroke Rehabil 2017 Jan;24(1):68-79. [doi: 10.1080/10749357.2016.1192361] [Medline: 27309680]
44. Rajesh Desai P, Nikhil Desai P, Deepak Ajmera K, Mehta K. A review paper on Oculus Rift-a virtual reality headset. Int J Eng Trends Technol 2014 Jul 25;13(4):175-179. [doi: 10.14445/22315381/ijett-v13p237]
45. McCue M, Fairman A, Pramuka M. Enhancing quality of life through telerehabilitation. Phys Med Rehabil Clin N Am 2010 Feb;21(1):195-205. [doi: 10.1016/j.pmr.2009.07.005] [Medline: 19951786]

XSL•FO
RenderX

46. Agostini M, Moja L, Banzi R, Pistotti V, Tonin P, Venneri A, et al. Telerehabilitation and recovery of motor function: a systematic review and meta-analysis. J Telemed Telecare 2015 Jun;21(4):202-213. [doi: 10.1177/1357633X15572201] [Medline: 25712109]
47. Gregory P, Alexander J, Satinsky J. Clinical telerehabilitation: applications for physiatrists. PM R 2011 Jul;3(7):647-56; quiz 656. [doi: 10.1016/j.pmrj.2011.02.024] [Medline: 21777864]
48. Hailey D, Roine R, Ohinmaa A, Dennett L. Evidence of benefit from telerehabilitation in routine care: a systematic review. J Telemed Telecare 2011;17(6):281-287. [doi: 10.1258/jtt.2011.101208] [Medline: 21844172]
49. Feys P, Coninx K, Kerkhofs L, De Weyer T, Truyens V, Maris A, et al. Robot-supported upper limb training in a virtual learning environment : a pilot randomized controlled trial in persons with MS. J Neuroeng Rehabil 2015 Jul 23;12:60 [FREE Full text] [doi: 10.1186/s12984-015-0043-3] [Medline: 26202325]
50. Maris A, Coninx K, Seelen H, Truyens V, De Weyer T, Geers R, et al. The impact of robot-mediated adaptive I-TRAVLE training on impaired upper limb function in chronic stroke and multiple sclerosis. Disabil Rehabil Assist Technol 2018 Jan;13(1):1-9. [doi: 10.1080/17483107.2016.1278467] [Medline: 28125300]
51. Casuso-Holgado M, Martín-Valero R, Carazo A, Medrano-Sánchez EM, Cortés-Vega MD, Montero-Bancalero F. Effectiveness of virtual reality training for balance and gait rehabilitation in people with multiple sclerosis: a systematic review and meta-analysis. Clin Rehabil 2018 Sep;32(9):1220-1234. [doi: 10.1177/0269215518768084] [Medline: 29651873]
52. Baram Y, Miller A. Virtual reality cues for improvement of gait in patients with multiple sclerosis. Neurology 2006 Jan 24;66(2):178-181. [doi: 10.1212/01.wnl.0000194255.82542.6b] [Medline: 16434649]
53. Sampson P, Freeman C, Coote S, Demain S, Feys P, Meadmore K, et al. Using functional electrical stimulation mediated by iterative learning control and robotics to improve arm movement for people with multiple sclerosis. IEEE Trans Neural Syst Rehabil Eng 2016 Feb;24(2):235-248. [doi: 10.1109/TNSRE.2015.2413906] [Medline: 25823038]
54. Kalron A, Achiron A, Pau M, Cocco E. The effect of a telerehabilitation virtual reality intervention on functional upper limb activities in people with multiple sclerosis: a study protocol for the TEAMS pilot randomized controlled trial. Trials 2020 Aug 12;21(1):713 [FREE Full text] [doi: 10.1186/s13063-020-04650-2] [Medline: 32787896]
55. Virtual reality approach in multiple sclerosis. ClinicalTrials.gov. URL: https://www.clinicaltrials.gov/ct2/show/NCT04212689 [accessed 2019-12-27]

Abbreviations

AE: adverse event
MS: multiple sclerosis
QoL: quality of life
RCT: randomized controlled trial
ROM: range of motion
TEAMS: Telerehabilitation and Multiple Sclerosis
VR: virtual reality

Edited by N Zary; submitted 15.01.22; peer-reviewed by M Pau, C Li; comments to author 27.05.22; revised version received 21.06.22; accepted 11.08.22; published 12.09.22.

Please cite as:
Kalron A, Frid L, Fonkatz I, Menascu S, Dolev M, Magalashvili D, Achiron A
The Design, Development, and Testing of a Virtual Reality Device for Upper Limb Training in People With Multiple Sclerosis: Single-Center Feasibility Study
JMIR Serious Games 2022;10(3):e36288
URL: https://games.jmir.org/2022/3/e36288
doi:10.2196/36288
PMID:36094809

©Alon Kalron, Lior Frid, Iliya Fonkatz, Shay Menascu, Mark Dolev, David Magalashvili, Anat Achiron. Originally published in JMIR Serious Games (https://games.jmir.org), 12.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality–Based Rehabilitation as a Feasible and Engaging Tool for the Management of Chronic Poststroke Upper-Extremity Function Recovery: Randomized Controlled Trial

Alejandro Hernandez[1], BEng, MSc; Liudmila Bubyr[2], BSc, MSc; Philippe S Archambault[1,3], BSc, OTR, PhD; Johanne Higgins[1,4], BSc, OTR, PhD; Mindy F Levin[1,3], BSc, PT, PhD; Dahlia Kairy[1,4], BSc, PT, PhD

[1]Centre for Interdisciplinary Research in Rehabilitation, Montreal, QC, Canada

[2]Independent biostatistician, Montreal, QC, Canada

[3]School of Physical & Occupational Therapy, McGill University, Montreal, QC, Canada

[4]Ecole de sciences de la réadaptation, Université de Montréal, Montreal, QC, Canada

Corresponding Author:
Alejandro Hernandez, BEng, MSc
Centre for Interdisciplinary Research in Rehabilitation
6363 Hudson Road
Office 061, Lindsay Pavilion of the Institut universitaire sur la réadaptation en déficience physique de Montréal
Montreal, QC, H3S 1M9
Canada
Phone: 1 514 737 0592 ext 4778
Email: alejandro.hernandez@mail.mcgill.ca

Abstract

Background: A growing number of stroke survivors are left with little to no rehabilitation services upon discharge from stroke rehabilitation, although arm deficits may persist or develop from disuse once rehabilitation services have ceased. Virtual reality (VR)–based rehabilitation, combined with new technologies such as telerehabilitation, including serious games using VR environments that encourage users to practice functional movements from home with minimal supervision, may have an important role to play in optimizing and maintaining upper extremity (UE) function.

Objective: The primary objective of this study is to determine the extent to which a 1-month intervention using a VR-based serious game is effective in improving UE function compared with an evidence-based home exercise program. A secondary objective is to assess the feasibility of implementing the intervention for chronic stroke rehabilitation in participants' homes.

Methods: A total of 51 chronic stroke participants were randomized to treatment (n=26, 51%; Jintronix system) or standard care (n=25, 49%; standardized Graded Repetitive Arm Supplementary Program kit home program) groups. The participants were evaluated at baseline (before), immediately after the intervention (after), and at follow-up (4 weeks). The primary outcome measure was the Fugl-Meyer Assessment for UE (FMA-UE). Secondary outcome measures included the Stroke Impact Scale and an abridged version of the Motor Activity Log-14. Self-reported number of sessions was logged for the standard care group.

Results: No statistically significant differences between groups were found across measures. Overall time effects were found for the FMA-UE (P=.045), specifically between preintervention and postintervention time points for both groups (P=.03). A total of 9 participants in the treatment group reached or surpassed the minimal clinically important difference in scores for the FMA-UE, with 7 (78%) of them having baseline low or moderate arm function, compared with 3 (33%) participants in the standard care group. Furthermore, 56% (9/16) of the participants in the treatment group who actively engaged with the system reached the minimal clinically important difference for the FMA-UE, compared with none for the 0% (0/10) less-active participants.

Conclusions: These findings suggest that UE training for chronic stroke survivors using virtual rehabilitation in their home may be as effective as a gold standard home exercise program and that those who used the system the most achieved the greatest improvement in UE function, indicating its relevance to being included as part of ongoing rehabilitation services.

Trial Registration: ClinicalTrials.gov NCT02491203; https://clinicaltrials.gov/ct2/show/NCT02491203

International Registered Report Identifier (IRRID): RR2-10.1016/j.cct.2015.12.006

(JMIR Serious Games 2022;10(3):e37506) doi:10.2196/37506

XSL•FO
RenderX

KEYWORDS

rehabilitation; serious game; stroke; telerehabilitation; upper extremity; virtual reality–based rehabilitation; virtual reality; virtual care

Introduction

Background

As of 2019, there were over 400,000 stroke survivors in Canada alone, a number that is projected to double by 2040 [1]. Hemiplegia, or weakness of one side of the body, can often translate into loss of upper extremity (UE) function. Unfortunately, the rate of full recovery of the affected arm was found to be only approximately 40%, especially in more severe cases [2]. Furthermore, recent trends in health care delivery often result in a shorter length of stay for an increasing number of stroke survivors, in spite of persistent functional deficits [3]. Outpatient or home care services may provide some rehabilitation care for a short time after stroke, but they are limited by long distances to and from home, high travel costs, and limited availability of caregivers [4].

The Canadian Stroke Best Practices Recommendations, updated in 2019, provide guidance for the provision of rehabilitation services [1]. On the basis of the evaluation of a stroke survivor's arm function, early treatment and individualized therapies of appropriate intensity and duration are recommended to optimize recovery in an inpatient clinical setting or on an outpatient basis, including during the chronic phase of stroke recovery. Providing intensive, meaningful, task-specific exercises to restore sensorimotor function is an important component of the recommended rehabilitation interventions, including traditional as well as more recent approaches, such as constraint-induced movement therapy and virtual reality (VR). VR provides an opportunity for the person to engage in repetitive movements and has been recognized as a valid complement to standard therapy [1].

The potential benefits of applying VR technology in physical rehabilitation notwithstanding, it is still unclear what therapy dosage levels are required to achieve optimal recovery, especially when considering factors such as time since stroke and severity of motor deficits. Dosage can be measured according to 3 distinct parameters: frequency, or number of exercise sessions per week; duration, or the period over which therapy is delivered; and number of repetitions, or time spent in active therapy, with an emphasis on the practice of challenging rather than overlearned tasks [5]. A study using motor learning methods as experimental interventions, observed statistically significant improvements in arm function after 300 hours of arm therapy practice over a 12-week period [6]. Furthermore, it would seem that the right combination of sufficiently high dosage and intensity training may be key to maintaining UE gains over the long term [7].

The provision of remote rehabilitation solutions has varied greatly over the years in their levels of technological sophistication, whether through simple telephone communications or more complex videoconferencing solutions and finally toward the more recent development of sensor and remote monitoring technologies that enable web-based applications to be deployed in the home [8]. Telehealth is an emerging technology that enables remote communication between patients and health professionals across health care fields, such as physical rehabilitation [9]. Communication can occur in real time through secure web-based platforms, allowing for face-to-face meetings. It may also occur asynchronously with therapists and patients logging onto platforms at different times to exchange relevant information. Telerehabilitation—or telehealth in the context of rehabilitation—could increase access to rehabilitation services by allowing for the remote supervision of patients who would otherwise be ineligible for or unable to access rehabilitation services following discharge. A systematic review of studies on motor recovery after a stroke suggests that interventions via telerehabilitation can be as effective as conventional in-person therapy [10].

As a complement to telerehabilitation, novel clinically oriented video gaming consoles, often referred to as serious games, are becoming increasingly accessible in health care settings. VR-based rehabilitation is increasingly accepted in the clinical setting for engaging patients to perform exercises and tasks repeated many times, which is the main principle of practice in standard care after stroke UE rehabilitation [11]. Older, widely available commercial platforms such as the Nintendo Wii gaming console were designed to physically engage the user in sport-like activities. A literature review examining the feasibility and effectiveness of commercial gaming consoles found that all 10 studies using Nintendo Wii as an intervention program showed gains in functional UE measures [12]. This promising finding suggests that Nintendo Wii and other similar VR-based serious games could be used to support recovery efforts in the clinical setting, although it remains to be verified whether their role can go beyond serving as adjuncts to standard therapy [13]. Not all VR serious game systems are equipped with a telehealth feature enabling participants to communicate remotely with a therapist; those that do provide a unique opportunity to customize rehabilitation interventions.

The Jintronix (Jintronix Inc) gaming console was designed to engage stroke survivors to recover lost UE function through a series of interactive games that encourage repetitive arm movements. A pilot study using the Jintronix system in a 2-arm randomized clinical trial with an acute poststroke clientele concluded that the VR gaming console was safe and feasible in its capacity to complement traditional therapy [14]. A meta-analysis concluded that similar home-based telerehabilitation approaches were as feasible as usual care [15]. Other studies using home-based interventions have reported modest UE gains in chronic poststroke clientele [16,17].

Objectives

First conceived as a support tool for stroke survivors, the Jintronix system presents itself as a promising tool to allow poststroke patients to pursue their UE rehabilitation, but who are no longer receiving standard care in the months or years since their discharge. At the onset of our study, no previous

XSL•FO
RenderX

studies have investigated the use of the Jintronix system as a remotely supervised home-based program for UE rehabilitation in chronic poststroke clientele. Therefore, the primary aim of this study was to assess the efficacy of a month-long home-based Jintronix system intervention in promoting UE functional recovery in chronic poststroke patients no longer receiving rehabilitation services. In line with the home-based nature of the intervention, as a secondary aim, we examined the feasibility of implementing the system in the homes of chronic stroke survivors.

Methods

Study Design

A single-blind (evaluator-blinded) parallel, 2-arm randomized controlled trial with a before, after, and follow-up design was used for this study in a chronic stroke population [18].

Ethical Considerations

This study was granted ethics board approval by the Research Ethics Board of the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR-937-0214). All participants provided informed consent before participating in the study.

Recruitment

A block randomization strategy with a block size of 6 using a random number generator was carried out by the study coordinator to randomly allocate participants into 1 of 2 distinct intervention groups. Sealed envelopes containing the group's identity were sequentially numbered according to initial randomization by block order. Allocation was performed previously but only revealed following the first in-person evaluation by the study coordinator (participants could not be blinded to the group assignation). Each intervention consisted of a 4-week long program, which was broken down as follows:

1. Treatment: home-based exercise program via the Jintronix system monitored offline by a therapist.
2. Standard care: home-based exercise program manual (Graded Repetitive Arm Supplementary Program [GRASP]) provided by a therapist without further supervision.

Textbox 1 outlines the participation inclusion and exclusion criteria.

Textbox 1. Participant inclusion and exclusion criteria.

Inclusion criteria

- First-time stroke having occurred >6 months earlier
- Having residual mild to moderate upper extremity (UE) impairments with a 2 to 6 score on the Chedoke-McMaster arm component (a quick screening tool used to ensure an adequate level of movement for the program) [19]
- No longer receiving rehabilitation services

Exclusion criteria

- Insufficient motor control to move the avatar onscreen
- Visual or auditory deficit
- Inability to understand simple verbal instructions
- Insufficient sitting balance
- Shoulder pain or pre-existing UE impairment limiting arm movement

Home-Based Intervention

The Jintronix system—composed of the Jintronix software installed on a computer, a large screen, and a Microsoft Kinect depth-detecting infrared sensor camera—was connected to the web via the participant's internet service provider, or through an internet key we provided if no service was installed. The Kinect camera tracks limb movements within a 3-meter range in 3D space without the need for a handheld controller. Data extracted by the camera are transferred in real time to the Jintronix software, which outputs a display of an avatar onscreen reflecting the user's movements. For example, the kitchen activity invites the user to reach a target placed in a virtual 3D kitchen setting; another activity requires bilateral movements of the arms to catch, carry, and drop objects in a 2D plane (Figure 1). The purpose of the system was to engage participants in repeated unilateral and bilateral UE movements to achieve satisfactory game scores needed to progress through the difficulty levels. User performance statistics such as movement speed and accuracy as well as overall game score could be accessed from the company servers.

XSL•FO
RenderX

Figure 1. The Jintronix system. Clockwise from top left: Fish Frenzy, Catch-Carry-Drop, Kitchen, Pop Clap game activities.



Participants randomized into the treatment group first set up a short meeting with a trained physiotherapist who provided them with a brief tutorial on the use of the Jintronix system and later set up a time with a technician to schedule the system's home installation. Participants were recommended to follow the program 5 times a week for ≥20 minutes per session. Taking into account the baseline level of UE function, the therapist custom tailored a simple program for each participant. Program progress was remotely monitored through the Jintronix system once to twice a week, and the level of difficulty, speed, and trajectories of arm movements were remotely and asynchronously adjusted by the therapist to maintain an optimal challenge according to the participant's UE improvement throughout the intervention.

The standard care group participants were provided with a manual for a standardized exercise program, the GRASP. The GRASP has been found to be effective as a supplement to ongoing UE rehabilitation during subacute stroke [20] or as a treatment alternative for discharged chronic stroke patients [21]. No therapists supervised participants' progress in the standard care group. A meeting was arranged with each participant before beginning of intervention to cover the components of the program. Participants were encouraged to engage in the program as many times a week as possible. Participants self-reported the number of sessions completed at the end of the 4 weeks. Although the GRASP program may have shared some similarities with the treatment (Jintronix system) program, such as promoting a variety of movements of the elbow joint, it also promoted wrist joint and hand dexterity exercises that were not included in the treatment program. Above all, a significant difference between the 2 groups lay in the provision of the program content: the treatment group program provided a very interactive visual and auditory experience, whereas the standard care group program provided the user with a simple manual in booklet form, as is often provided in outpatient care.

Outcome Measures

Baseline demographics were collected for all participants on their first visit. A total of 3 clinical efficacy outcome measures were selected to assess functional changes in the upper limb. The primary clinical outcome measure consisted of the Fugl-Meyer Assessment for UE (FMA-UE), which quantifies UE impairment after stroke. A gold standard in clinical practice, it has high interrater reliability and content validity [22] and is widely used across a range of clinical studies targeting poststroke recovery [23]. The Stroke Impact Scale (SIS; SIS 3.0; along with its individual Strength, activities of daily living [ADL], Mobility, and Hand Function components) and the Motor Activity Log (MAL; MAL-14, abridged 14-question version) were used as secondary outcome measures to self-assess quality of life and the use of the impaired arm in ADL, respectively. Both outcomes were chosen for their strong internal consistency and test-retest reliability [24,25].

To assess the feasibility of implementing the Jintronix system at home, several variables were collected, among which are the following: time of home installation, number of sessions throughout the 4-week intervention period, total time spent on the program, pain and fatigue indicators, and episodes of dizziness or falls.

Evaluations using these clinical outcome measures were carried out at baseline (before), after the intervention (after), and at 4-week follow-up (follow-up). Evaluators were blinded to participant group allocation and were not involved in the interventions.

XSL•FO
RenderX

Statistical Analysis

Overview

Demographic variables exhibiting normal data distribution were represented by their means and SDs. Median and IQRs were used when describing data that were not normally distributed. Unlike the mean, the median is more robust against the effect of potential outliers in overall as well as subgroup analyses [26]. Effects were tested against a significance level of Cronbach α=.05. When available, the minimal clinically important difference (MCID) was used as a cutoff to determine clinically meaningful differences.

Normality of outcome distributions was assessed using the Kolmogorov-Smirnov test (P>.05 for normality) to investigate further differences in the groups. Either 2-sample t tests or Wilcoxon 2-sample tests were conducted depending on the normal or nonparametric nature of the distributions, respectively.

Clinical data collected from onsite assessments as well as data recorded from the Jintronix system were stored in a secure database REDCap (Research Electronic Data Capture; Vanderbilt University). Statistical analyses were carried out using the SAS 9.4 (SAS Institute) software package.

Sample Size

A sample size of 26 participants per group was determined using G*Power, assuming a medium effect size of 0.2, accounting for a 20% attrition rate, and setting the Cronbach α to .05 and the power to 0.8.

Modeling

A mixed model paradigm with a compound symmetry structure was used to model the analyses. The group and time variables were set as factors to account for between-participant and within-participant differences. An adjustment for baseline differences in variables such as participant age and arm function at the onset of participation was performed to correct for improprieties in baseline characteristics from the observed data, thus making the compared groups more homogeneous.

Subgrouping Analyses

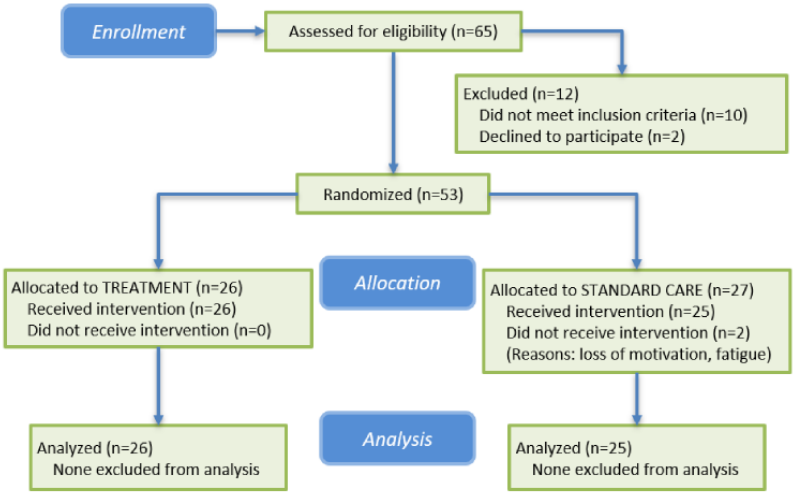
The *baseline FMA-UE score* was chosen as a subgroup factor to further explore its role in participant improvement across outcome measures. The FMA-UE cutoff scores were used to define the factor's 3 levels: low, moderate, and high function. The choice of cutoff scores was based on previous studies examining the FMA-UE as a factor [27,28]. Similarly, to further explore efficacy, *active playing time* was used to create a 2-level factor to categorize treatment group participants around the 400-minute cutoff time, as recommendations were for participants to engage in five 20-minute sessions per week (totaling 400 minutes for the entirety of the 4-week program).

Results

Participant Demographics

A total of 53 chronic poststroke individuals consented to participate in this study. Of the 53 participants, 2 (4%) withdrew following consent and randomization into the standard care group, with reasons cited being loss of motivation or fatigue; 51 participants completed the study (n=26, 51% and n=25, 49% for the treatment and standard care groups, respectively). A participant enrollment flow diagram is presented in Figure 2.

Figure 2. Flow diagram of the study's enrollment process.



In total, 27% (14/51) of the participants were female, with group ratios differing slightly (9/26, 35% for treatment and 5/25, 20% for standard care). Mean participant age was 59.8 (SD 13.1) years for treatment and 56.7 (SD 11.2) years for standard care. Median time since stroke was 63 months (IQR 5.3 years) and 53 months (IQR 4.4 years) for the treatment and standard care groups, respectively. There were no statistically significant differences between groups for the list of relevant participant demographics, as outlined in Table 1.

XSL•FO
RenderX

Table 1. Participant demographics at baseline (before the intervention).

Variables	Treatment (n=26)	Standard care (n=25)
Age (years), mean (SD)	59.8 (13.1)	56.7 (11.2)
Male, n (%)	17 (65)	20 (80)
Stroke type, n (%)		
Ischemic	14 (54)	10 (40)
Hemorrhagic	7 (27)	7 (28)
Unknown	5 (19)	8 (32)
Handedness, n (%)		
Left	2 (8)	3 (12)
Right	23 (88)	22 (88)
Ambidextrous	1 (4)	0 (0)
Left-side hemiparesis, n (%)	13 (50)	11 (44)
Dominant side affected, n (%)	15 (58)	13 (52)
Time since stroke, median (IQR)		
In years	5.3 (1.5-8.1)	4.4 (2.2-7.4)
In months	63 (18-97)	53 (26-89)
Montreal Cognitive Assessment score, median (IQR)	25 (20-27)	25 (24-27)
Chedoke-McMaster score, median (IQR)	4 (3-5)	4 (3-5)
Fugl-Meyer Assessment for upper-extremity score, median (IQR)	30 (17-52)	38 (22-55)

Overall Group Analyses

Mixed model analysis (adjusted for baseline differences) revealed no overall statistically significant differences between groups across all outcome measures (Table 2). However, there was a significant time effect for the FMA-UE (P=.046) and SIS-total (P=.048) outcome measures. In particular, for the FMA-UE, a significant time effect was observed in the before to after periods (P=.03) but not for the other periods, including between after and follow-up (and before and follow-up). Although the FMA-UE trended toward better scores between before and after periods (P=.08), no group-by-time interactions were found to be statistically significant across any of the measures.

Table 2. Mixed models results across outcome measures by effect type.

Outcome measure	Group		Time		Group×time	
	*F*test (*df*)	*P* value	*F*test (*df*)	*P* value	*F*test (*df*)	*P* value
Fugl-Meyer Assessment for upper extremity	1.50 (34)	.23	3.19 (86)	.046	2.62 (86)	.08
Motor Activity Log						
Amount	0 (34)	.98	.30 (86)	.74	.17 (86)	.85
Quality	0 (34)	.99	1.00 (86)	.37	.03 (86)	.97
Stroke Impact Scale						
Strength	1.81 (32)	.19	1.89 (80)	.16	0.04 (80)	.96
Activities of daily living	0.84 (32)	.37	0.94 (78)	.40	2.42 (78)	.09
Mobility	0.80 (34)	.38	1.67 (85)	.19	1.52 (85)	.22
Hand function	0.48 (34)	.49	0.94 (86)	.39	0.80 (86)	.45
Total	0.09 (27)	.76	3.17 (72)	.048	2.14 (72)	.12

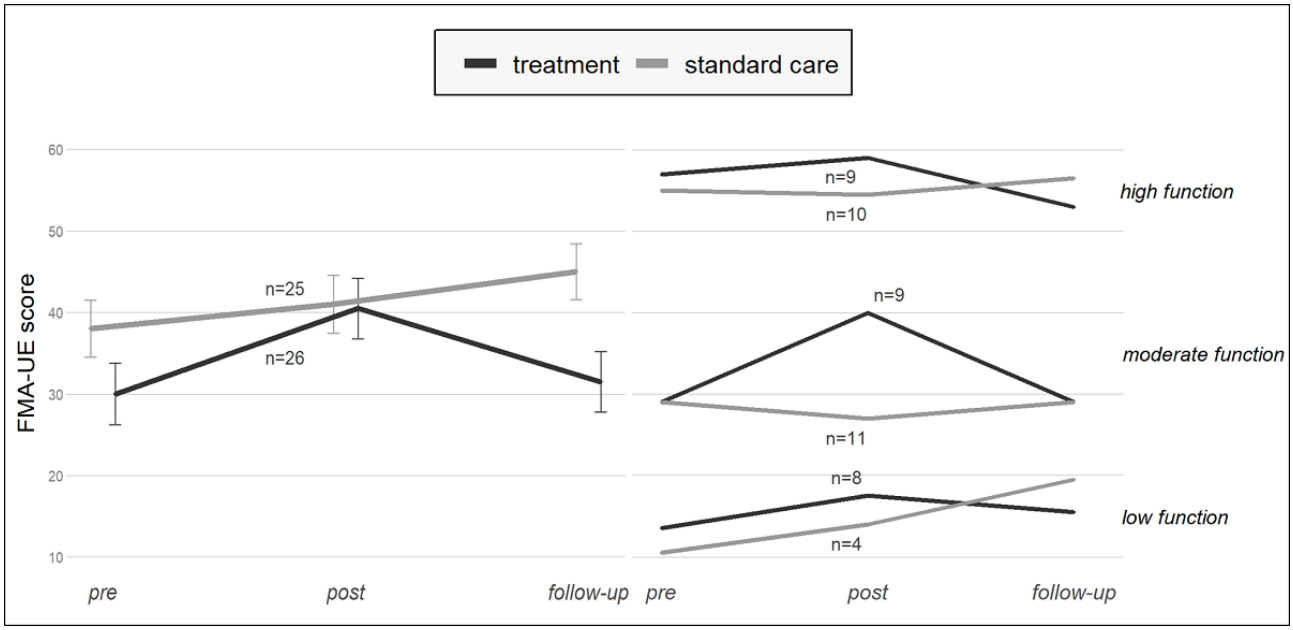
Outcome measure distributions were mostly observed to exhibit nonnormal distributions, with the exceptions of SIS-ADL and SIS-mobility, which were normally distributed (P=.15 and P=.12, respectively). The 2-sample Wilcoxon tests revealed no statistically significant differences between groups across time points (before, after, and follow-up). No significant differences were observed in the results of the 2-sample t tests carried out on the SIS-ADL and SIS-mobility measures.

The groups' median FMA-UE scores over time are shown in Figure 3 (left-hand side). The upward trend between the before

XSL•FO
RenderX

and after time points tends to corroborate the significant time effects identified. Gains obtained by the treatment group after the intervention were no longer seen at follow-up, as seen in Figure 3 (right-hand side).

Figure 3. Left: median Fugl-Meyer Assessment for upper-extremity (FMA-UE) score over time by group; right: by group and baseline arm function.



Subgroup Analysis by Baseline FMA-UE

Mixed model analyses using baseline FMA-UE as a subgroup factor revealed a statistically significant difference between groups for the SIS-strength measure, but only for the high-function subcohort (*P*=.046). Post hoc analyses of SIS-strength scores after the intervention revealed a significant difference between the groups (*P*=.008). However, no statistical differences were found at the follow-up. No other SIS component measure (ADL, mobility, or hand function) produced significant differences between groups or within the subgroup analyses.

The groups' median FMA-UE scores per baseline FMA-UE are shown in Figure 3 (right side). Although the differences were not statistically significant, an 11-point change in the median FMA-UE score was observed before and after the intervention for the treatment with the moderate FMA-UE subgroup (n=9 participants). This was the only subgroup that surpassed the 5-point MCID for the FMA-UE measure.

The number and proportion of participants across both groups that either reached or surpassed the MCID threshold (positive 5-point FMA-UE change in score) are displayed in Table 3. Nearly half of the participants (7/17, 41%) in the low- and moderate-function subcohort from the treatment group reached the MCID, whereas approximately a fifth of the participants (3/15, 20%) were observed for the same subcohort from the standard care group. Overall, 35% (9/26) of the treatment group participants achieved scores at or above the MCID, a little under double the ratio seen in standard care participants (5/25, 20%), although a chi-square analysis did not support these rates as statistically significant (*P*=.32).

Table 3. Number of participants having reached or surpassed the minimal clinically important difference (MCID) on the Fugl-Meyer Assessment for upper extremity (FMA-UE) after intervention, according to group and baseline arm function. Number and proportion of treatment group participants having reached or surpassed the MCID on the FMA-UE after the intervention, by levels of gameplay time and baseline arm function.

Group	Baseline FMA-UE arm function level, MCID/n[a], %[b]			Totals, MCID/n, %
	Low	Moderate	High	
Treatment	3/8, 38	4/9, 44	2/9, 22	9/26, 35
Standard care	1/4, 25	2/11, 18	2/10, 20	5/25, 20
<400 minutes	0/4, 0	0/2, 0	0/4, 0	0/10, 0
>400 minutes	3/4, 75	4/7, 57	2/5, 40	9/16, 56

[a]MCID/n: ratio of participants reaching MCID on total subgroup number.

[b]Ratio percentage.

Intervention Group Feasibility and Efficacy

The key descriptive feasibility and efficacy findings for the treatment group participants are shown in Table 4. The standard care group participants engaged in the GRASP program for a median of 12 (self-reported) sessions over 4 weeks, with 50% (11/22) of participants ranging between 8 and 16 sessions. Participants in the treatment group engaged with the Jintronix

XSL•FO
RenderX

system for a median of 21.5 sessions and invested a total duration of 527 minutes (with 13/26, 50% of participants ranging between 310 and 673 minutes). Of particular note were participants in the treatment group with moderate arm function: they tended to spend more time exercising (median 652 minutes) compared with the low and high functional participants. In addition, the more active participants gained a median of 5.5 points in their FMA-UE scores compared with 0 for the less-active participants and 1 for the standard care group.

Table 4. Treatment group participant statistics following a 4-week intervention.

	Population size, N	Number of sessions, median (IQR)[a]	Time (minutes), median (IQR)	Change in Fugl-Meyer Assessment for upper extremity, median (IQR)
Standard care	22[b]	12 (8 to 16)	N/A[c]	1 (−2 to 4)
Treatment	26	21.5 (16 to 27)	527 (310 to 673)	2 (0 to 6.8)
By Fugl-Meyer Assessment for upper-extremity level				
Low	8	19 (16 to 22)	431 (237 to 660)	3 (0.5 to 6.8)
Moderate	9	26 (22 to 30)	652 (479 to 864)	2 (0 to 9)
High	9	21 (13 to 27)	468 (287 to 570)	0 (−1 to 3)
By total duration				
<400 minutes	10	15 (11.5 to 16.8)	269 (152 to 317)	0 (−1 to 2.8)
>400 minutes	16	26.5 (22 to 30.5)	648 (561 to 855)	5.5 (0 to 9)

[a]IQR expressed as (25th percentile-75th percentile).

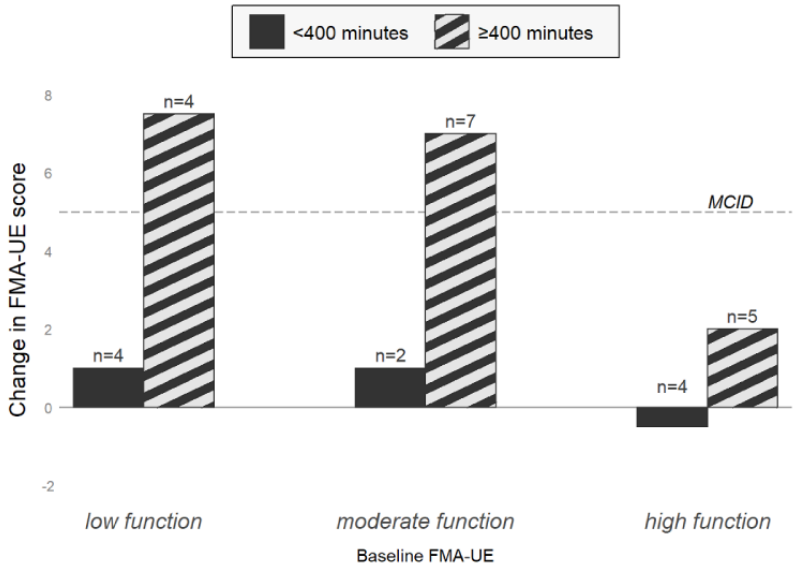
[b]Data available for 22 of the 25 standard care group participants.

[c]N/A: not applicable.

The installation time on the Jintronix system at home ranged between 15 and 40 minutes. Participants in the treatment group reported median fatigue and pain scores of 3.3 and 1.8, respectively, both rated on a 10-point scale (with 10 representing the maximum). No adverse events, such as falls or episodes of dizziness, were reported by any participant. Two participants reported difficulties with the technology, primarily related to controlling the mouse and navigating the gaming interface.

The median change in FMA-UE score after the intervention is shown along a continuum of baseline arm function levels (Figure 4). A downward trend can be observed in the FMA-UE gains as the baseline arm function increased for the more active subgroup (diagonal patterned bars). Of note are the participants whose arm function were either low or moderate; both subgroups achieved a median of 7 or higher increase in the FMA-UE (above the 5-point MCID). By contrast, the dark bars suggest that participants having invested less than the recommended dosage tended to produce little or no gains regardless of baseline arm function. Despite the visual trends displayed in Figure 4, the differences were not statistically significant (P=.08 for the low function subgroup and P=.55 and P=.27 for the moderate and high subgroups, respectively).

Figure 4. Change in median Fugl-Meyer Assessment for upper-extremity (FMA-UE) score after intervention for treatment group participants; by baseline arm function and level of gameplay duration. MCID: minimal clinically important difference.



XSL•FO
RenderX

Interestingly, none of the 10 less-active participants achieved the 5-point FMA-UE, which indicates a clinically important change, whereas 56% (9/16) of more active participants achieved significant gains in arm function (Table 3).

Discussion

Principal Findings

Within the context of this randomized controlled trial, we examined the efficacy and feasibility of a noninvasive VR-based rehabilitation serious game for UE training with a chronic poststroke clientele no longer receiving rehabilitation services. Participants in both the treatment and standard care groups were able to successfully engage in their assigned interventions. After accounting for baseline differences across participant characteristics, no statistically significant differences were found between groups across all outcome measures. However, both groups showed statistically significant improvements in the FMA-UE and SIS outcome measures over time, particularly between the periods before and after the interventions. Therefore, the program based on the Jintronix system was noninferior compared with the standardized GRASP program in improving UE function. All participants remained within the group to which they were assigned, thereby respecting the intention-to-treat principle, although not all participants achieved the recommended dosage. Therefore, further subgroup analyses were conducted to better understand the dosage administered to the treatment group.

Amount of Time Played Makes a Difference

The more active treatment group participants improved on their FMA-UE scores by a median of 5.5 points, whereas their less-active counterparts gained a 0-point median change in score. Treatment group participants lost most gains acquired on the month following the end of intervention. Similar observations were noted in studies conducting UE treatment programs with chronic poststroke patients [29,30].

Findings from within the treatment group suggest that participants having invested more time engaging in the activity (Figure 4) seem to confirm that “the more one puts into one’s recovery, the more one gets out of it” [31]. This appears to hold true for both low and moderate FMA-UE subcohorts but less so for the high FMA-UE subcohort.

In addition to the amount of time spent exercising, there were important differences between programs that included more games conducive to repetitive movements of the shoulder and elbow joints by the Jintronix system, whereas the GRASP program included a considerable focus on wrist and finger movements. Therefore, the amount of time cannot be isolated from the rest of the intervention itself. However, the results of this study support those of previous studies which found that game-based rehabilitation systems could spark a greater interest in the participant, which could make it easier to spend more time on a program than usual care [32].

Baseline Function Plays a Role in the Rate of Recovery

Subgrouping participants by level of arm function showed a trend toward differences in group scores in the moderate FMA-UE cohort (Figure 3), with an 11-point median FMA-UE score difference between groups (statistically nonsignificant). The data observed suggest minimal clinically meaningful changes in the treatment group. However, higher-functioning poststroke participants may have greater difficulty in obtaining greater gains, in part perhaps due to the ceiling effect of the outcome measure. A better understanding of the relationship between high function and the extent of improvement would be worthy of further exploration to best determine dosage. In spite of the clinical measure’s demonstrated content validity and reliability, the baseline FMA-UE score may be less responsive to change when it is already high to begin with [33]. Participants with low baseline FMA-UE had smaller gains than moderately functional stroke survivors but greater gains than the higher-functioning participants.

Kinematic measures could fill in the gaps where established clinical measures fail to detect changes. Rather than only quantifying functional improvement using a MCID threshold, a more nuanced approach could be envisioned for participants at the higher and lower ends of the UE functional spectrum. This could be implemented via wearable sensors or robot-mediated consoles, which measure variables such as the speed of movement, range of motion, and path smoothness [34]. Some studies have found significant correlations between kinematic measures and the FMA-UE, although they caution against substituting out established clinical measures [35]. A meta-analysis concluded that kinematic measures can be good complements to clinical outcome measures as they are apt for detecting smaller improvements [36]. Although kinematic variables were not collected (given that they were outside the scope of this study’s primary objectives), they may provide added value to future studies aimed at implementing similar technologies as an adjunct to clinical outcome evaluations.

Sensor technology may become more omnipresent in the future, tracking arm activity to accurately account for activity metrics performed within and beyond a prescribed intervention program [37]. It could also serve in the collection of kinematic measures during participant evaluation, especially if done remotely or when established clinical outcome measures fail to detect smaller changes in arm function. In fact, its hands-free simplicity of use prompted a study to verify and confirm its validity as a means of assessing UE function in a clinical setting [38].

On the Question of Dosage

We provided simple participation guidelines formulated in such ways as “5 times per week, 20 minutes per session for 4 weeks,” in an effort to promote program engagement. This was based on prior examples of telerehabilitation intervention programs for upper-limb recovery after stroke [39,40]. A meta-analysis found that exercise dosage strongly predicted functional motor recovery when it was modeled as a linear regression of key predictor variables, such as dosage time and time since stroke [41]. This finding was also confirmed by a study that observed a linear relationship between dosage and functional outcome gains, but only up to a certain number of hours, beyond which the returns for any additional time tapered off [5].

Although on the one hand, our results showed nonretention of upper-limb gains by follow-up, on the other hand, it has been

XSL•FO
RenderX

suggested that task-specific repeated practice regimens induce lasting motor cortical reorganization that often precedes motor improvement [42]. We may not yet be sure of the complex interplay between exercise frequency, intensity, and duration needed to optimize recovery, but the results would suggest that extending treatment duration could allow sufficient time for motor cortical reorganization to make way for motor recovery of the upper limb.

Virtual Rehabilitation as an Additional Tool in the Management of Chronic Poststroke Upper-Extremity Recovery

Home-based intervention programs have been used in prior studies [43] as a central component of chronic poststroke study design. In this study, we included a control group that received an evidence-based standardized exercise program targeting the repetition of upper-limb movements that emulate ADL, which is currently frequently provided in rehabilitation programs. As such, participants in both groups benefited from a program that allowed for comparable levels of upper-limb activity.

While treatment group participants required a home installation of the Jintronix system, the setup was relatively simple, requiring minimal space in the participant's living spaces and little to no technical maintenance throughout the duration of the intervention. A certain degree of computer literacy was required of the participants to navigate the interface, which was addressed during the first meeting. Furthermore, the system had the added capacity to inform clinicians of participant progress and time spent on the activities, factors which appeared to play an important role in recovery.

The treatment group participants played for a median of 527 minutes of activity (approximately 9 hours). Most participants needed no extrinsic prodding to engage in the program, perhaps relying instead on their desire to engage in the visually rewarding gaming environment [32]. We would argue that the intuitiveness of VR game consoles facilitates self-directed behavior, which ultimately influences program engagement and adherence, in line with the positive connection to the gaming avatar participants reported in a Nintendo Wii environment [44]. On the basis of these considerations along with the technology's simplicity of use and installation, these findings support the feasibility of using VR serious games as tools for the management of chronic poststroke recovery, as recommended in the Guidelines for Adult Stroke Rehabilitation and Recovery [45].

Study Limitations and Future Directions

Although the outcome measures included had strong psychometric properties, there were some limitations. The FMA-UE measure may have limited ability to detect changes when participants are either low or high on baseline arm function. The SIS and MAL measures may lack sensitivity in detecting smaller changes in self-perceived function. In addition, the use of subgrouping strategies to compare differences in baseline FMA-UE scores limited the ability to detect significant effects given the small overall sample size.

Participants in the standard care group did not log specific time spent on the GRASP program. More precision could have been achieved had wearable sensor technology been available and integrated into the program, which could have accurately kept track of time spent on active movements of the upper limb.

Participants in both groups gained significant arm function improvements while they were actively engaged in 1 of the 2 programs. Rather than draw comparisons between groups, future studies could set out to evaluate novel programs against a standardized one from a perspective of noninferiority, so that clinicians have more tools available to choose from for UE rehabilitation.

Future studies could further explore the impact of extended treatment time and increased number of visits for a follow-up evaluation to more easily keep track of changes in arm function. This would allow the possibility to test the increasingly popular theory that an increase in dosage results in better functional outcomes.

Finally, artificial intelligence could eventually be incorporated into VR serious games to reduce clinician involvement by adjusting difficulty level, movement range, and type of activities based on the user's needs and preferences.

Conclusions

There were visible trends of improvement following intervention for both interventions, particularly when participants were most actively engaged with the system. Depending on the individual and clinical context, the results indicate that VR serious games with clinician monitoring may be additional, effective, and feasible tools to include in the long-term management of upper-limb rehabilitation after stroke.

Acknowledgments

This study was funded by the Canadian Institutes of Health Research. The views and opinions expressed are those of the authors and not necessarily those of the Canadian Institutes of Health Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 396 KB - games_v10i3e37506_app1.pdf]

XSL•FO
RenderX

References

1. Teasell R, Salbach NM, Foley N, Mountain A, Cameron JI, Jong AD, et al. Canadian stroke best practice recommendations: rehabilitation, recovery, and community participation following stroke. Part one: rehabilitation and recovery following stroke; 6th edition update 2019. Int J Stroke 2020 Oct;15(7):763-788. [doi: 10.1177/1747493019897843] [Medline: 31983296]
2. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. Stroke 2003 Sep;34(9):2181-2186. [doi: 10.1161/01.STR.0000087172.16305.CD] [Medline: 12907818]
3. Holden MK, Dyar TA, Dayan-Cimadoro L. Telerehabilitation using a virtual environment improves upper extremity function in patients with stroke. IEEE Trans Neural Syst Rehabil Eng 2007 Mar;15(1):36-42. [doi: 10.1109/TNSRE.2007.891388] [Medline: 17436874]
4. Sibley LM, Glazier RH. Reasons for self-reported unmet healthcare needs in Canada: a population-based provincial comparison. Healthc Policy 2009 Aug;5(1):87-101 [FREE Full text] [Medline: 20676253]
5. Lang CE, Lohse KR, Birkenmeier RL. Dose and timing in neurorehabilitation: prescribing motor therapy after stroke. Curr Opin Neurol 2015 Dec;28(6):549-555 [FREE Full text] [doi: 10.1097/WCO.0000000000000256] [Medline: 26402404]
6. McCabe J, Monkiewicz M, Holcomb J, Pundik S, Daly JJ. Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Jun;96(6):981-990 [FREE Full text] [doi: 10.1016/j.apmr.2014.10.022] [Medline: 25461822]
7. Ward NS, Brander F, Kelly K. Intensive upper limb neurorehabilitation in chronic stroke: outcomes from the Queen Square programme. J Neurol Neurosurg Psychiatry 2019 May;90(5):498-506. [doi: 10.1136/jnnp-2018-319954] [Medline: 30770457]
8. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform 2009;145:231-248. [Medline: 19592797]
9. Dodakian L, McKenzie AL, Le V, See J, Pearson-Fuhrhop K, Burke Quinlan E, et al. A home-based telerehabilitation program for patients with stroke. Neurorehabil Neural Repair 2017;31(10-11):923-933 [FREE Full text] [doi: 10.1177/1545968317733818] [Medline: 29072556]
10. Sarfo FS, Ulasavets U, Opare-Sem OK, Ovbiagele B. Tele-rehabilitation after stroke: an updated systematic review of the literature. J Stroke Cerebrovasc Dis 2018 Sep;27(9):2306-2318 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2018.05.013] [Medline: 29880211]
11. Bonnechère B, Jansen B, Omelina L, Van Sint Jan S. The use of commercial video games in rehabilitation: a systematic review. Int J Rehabil Res 2016 Dec;39(4):277-290. [doi: 10.1097/MRR.0000000000000190] [Medline: 27508968]
12. Pietrzak E, Cotea C, Pullman S. Using commercial video games for upper limb stroke rehabilitation: is this the way of the future? Top Stroke Rehabil 2014;21(2):152-162. [doi: 10.1310/tsr2102-152] [Medline: 24710975]
13. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
14. Norouzi-Gheidari N, Hernandez A, Archambault PS, Higgins J, Poissant L, Kairy D. Feasibility, safety and efficacy of a virtual reality exergame system to supplement upper extremity rehabilitation post-stroke: a pilot randomized clinical trial and proof of principle. Int J Environ Res Public Health 2019 Dec 23;17(1):113 [FREE Full text] [doi: 10.3390/ijerph17010113] [Medline: 31877910]
15. Huijgen BC, Vollenbroek-Hutten MM, Zampolini M, Opisso E, Bernabeu M, Van Nieuwenhoven J, et al. Feasibility of a home-based telerehabilitation system compared to usual care: arm/hand function in patients with stroke, traumatic brain injury and multiple sclerosis. J Telemed Telecare 2008;14(5):249-256. [doi: 10.1258/jtt.2008.080104] [Medline: 18633000]
16. Hsieh YW, Chang KC, Hung JW, Wu CY, Fu MH, Chen CC. Effects of home-based versus clinic-based rehabilitation combining mirror therapy and task-specific training for patients with stroke: a randomized crossover trial. Arch Phys Med Rehabil 2018 Dec;99(12):2399-2407. [doi: 10.1016/j.apmr.2018.03.017] [Medline: 29702070]
17. Langan J, Delave K, Phillips L, Pangilinan P, Brown SH. Home-based telerehabilitation shows improved upper limb function in adults with chronic stroke: a pilot study. J Rehabil Med 2013 Feb;45(2):217-220 [FREE Full text] [doi: 10.2340/16501977-1115] [Medline: 23319181]
18. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
19. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, et al. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. Stroke 1993 Jan;24(1):58-63. [doi: 10.1161/01.str.24.1.58] [Medline: 8418551]
20. Harris JE, Eng JJ, Miller WC, Dawson AS. A self-administered Graded Repetitive Arm Supplementary Program (GRASP) improves arm function during inpatient stroke rehabilitation: a multi-site randomized controlled trial. Stroke 2009 Jun;40(6):2123-2128. [doi: 10.1161/STROKEAHA.108.544585] [Medline: 19359633]
21. Rand D, Weingarden H, Weiss R, Yacoby A, Reif S, Malka R, et al. Self-training to improve UE function at the chronic stage post-stroke: a pilot randomized controlled trial. Disabil Rehabil 2017 Jul;39(15):1541-1548. [doi: 10.1080/09638288.2016.1239766] [Medline: 27793071]

XSL•FO
RenderX

22. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. Phys Ther 1993 Jul;73(7):447-454. [doi: 10.1093/ptj/73.7.447] [Medline: 8316578]
23. Veras M, Kairy D, Rogante M, Giacomozzi C, Saraiva S. Scoping review of outcome measures used in telerehabilitation and virtual reality for post-stroke rehabilitation. J Telemed Telecare 2017 Jul;23(6):567-587. [doi: 10.1177/1357633X16656235] [Medline: 27342850]
24. Uswatte G, Taub E, Morris D, Vignolo M, McCulloch K. Reliability and validity of the upper-extremity Motor Activity Log-14 for measuring real-world arm use. Stroke 2005 Nov;36(11):2493-2496. [doi: 10.1161/01.STR.0000185928.90848.2e] [Medline: 16224078]
25. Vellone E, Savini S, Fida R, Dickson V, Melkus G, Carod-Artal F, et al. Psychometric evaluation of the Stroke Impact Scale 3.0. J Cardiovasc Nurs 2015;30(3):229-241. [doi: 10.1097/JCN.0000000000000145] [Medline: 24695074]
26. McHugh ML. Descriptive statistics, part II: most commonly used descriptive statistics. J Spec Pediatr Nurs 2003;8(3):111-116. [doi: 10.1111/j.1088-145x.2003.00111.x] [Medline: 12942890]
27. Woodbury ML, Velozo CA, Richards LG, Duncan PW. Rasch analysis staging methodology to classify upper extremity movement impairment after stroke. Arch Phys Med Rehabil 2013 Aug;94(8):1527-1533. [doi: 10.1016/j.apmr.2013.03.007] [Medline: 23529144]
28. Hoonhorst MH, Nijland RH, van den Berg JS, Emmelot CH, Kollen BJ, Kwakkel G. How do Fugl-Meyer arm motor scores relate to dexterity according to the action research arm test at 6 months poststroke? Arch Phys Med Rehabil 2015 Oct;96(10):1845-1849. [doi: 10.1016/j.apmr.2015.06.009] [Medline: 26143054]
29. Llorens R, Noé E, Alcañiz M, Deutsch JE. Time since injury limits but does not prevent improvement and maintenance of gains in balance in chronic stroke. Brain Inj 2018;32(3):303-309. [doi: 10.1080/02699052.2017.1418905] [Medline: 29278927]
30. Hung JW, Chou CX, Chang YJ, Wu CY, Chang KC, Wu WC, et al. Comparison of Kinect2Scratch game-based training and therapist-based training for the improvement of upper extremity functions of patients with chronic stroke: a randomized controlled single-blinded trial. Eur J Phys Rehabil Med 2019 Oct;55(5):542-550 [FREE Full text] [doi: 10.23736/S1973-9087.19.05598-9] [Medline: 30781936]
31. Langhorne P, Wagenaar R, Partridge C. Physiotherapy after stroke: more is better? Physiother Res Int 1996;1(2):75-88. [doi: 10.1002/pri.6120010204] [Medline: 9238725]
32. Trombetta M, Bazzanello Henrique PP, Brum MR, Colussi EL, De Marchi AC, Rieder R. Motion Rehab AVE 3D: a VR-based exergame for post-stroke rehabilitation. Comput Methods Programs Biomed 2017 Nov;151:15-20. [doi: 10.1016/j.cmpb.2017.08.008] [Medline: 28946996]
33. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair 2002 Sep;16(3):232-240. [doi: 10.1177/154596802401105171] [Medline: 12234086]
34. Held JP, Klaassen B, Eenhoorn A, van Beijnum BJ, Buurke JH, Veltink PH, et al. Inertial sensor measurements of upper-limb kinematics in stroke patients in clinic and home environment. Front Bioeng Biotechnol 2018 Apr 12;6:27 [FREE Full text] [doi: 10.3389/fbioe.2018.00027] [Medline: 29707537]
35. Duret C, Courtial O, Grosmaire AG. Kinematic measures for upper limb motor assessment during robot-mediated training in patients with severe sub-acute stroke. Restor Neurol Neurosci 2016;34(2):237-245. [doi: 10.3233/RNN-150565] [Medline: 26890098]
36. Villepinte C, Verma A, Dimeglio C, De Boissezon X, Gasq D. Responsiveness of kinematic and clinical measures of upper-limb motor function after stroke: a systematic review and meta-analysis. Ann Phys Rehabil Med 2021 Mar;64(2):101366 [FREE Full text] [doi: 10.1016/j.rehab.2020.02.005] [Medline: 32156646]
37. Parker J, Powell L, Mawson S. Effectiveness of upper limb wearable technology for improving activity and participation in adult stroke survivors: systematic review. J Med Internet Res 2020 Jan 08;22(1):e15981 [FREE Full text] [doi: 10.2196/15981] [Medline: 31913131]
38. Kim WS, Cho S, Baek D, Bang H, Paik NJ. Upper extremity functional evaluation by Fugl-Meyer assessment scoring using depth-sensing camera in hemiplegic stroke patients. PLoS One 2016 Jul 1;11(7):e0158640 [FREE Full text] [doi: 10.1371/journal.pone.0158640] [Medline: 27367518]
39. Piron L, Turolla A, Agostini M, Zucconi C, Cortese F, Zampolini M, et al. Exercises for paretic upper limb after stroke: a combined virtual-reality and telemedicine approach. J Rehabil Med 2009 Nov;41(12):1016-1102 [FREE Full text] [doi: 10.2340/16501977-0459] [Medline: 19841835]
40. Henrique PP, Colussi EL, De Marchi AC. Effects of exergame on patients' balance and upper limb motor function after stroke: a randomized controlled trial. J Stroke Cerebrovasc Dis 2019 Aug;28(8):2351-2357. [doi: 10.1016/j.jstrokecerebrovasdis.2019.05.031] [Medline: 31204204]
41. Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. Stroke 2014 Jul;45(7):2053-2058 [FREE Full text] [doi: 10.1161/STROKEAHA.114.004695] [Medline: 24867924]
42. Page SJ, Gater DR, Bach-Y-Rita P. Reconsidering the motor recovery plateau in stroke rehabilitation. Arch Phys Med Rehabil 2004 Aug;85(8):1377-1381. [doi: 10.1016/j.apmr.2003.12.031] [Medline: 15295770]

XSL•FO
RenderX

43. Coupar F, Pollock A, Legg LA, Sackley C, van Vliet P. Home-based therapy programmes for upper limb functional recovery following stroke. Cochrane Database Syst Rev 2012 May 16;2012(5):CD006755 [FREE Full text] [doi: 10.1002/14651858.CD006755.pub2] [Medline: 22592715]
44. Tsekleves E, Paraskevopoulos IT, Warland A, Kilbride C. Development and preliminary evaluation of a novel low cost VR-based upper limb stroke rehabilitation platform using Wii technology. Disabil Rehabil Assist Technol 2016;11(5):413-422. [doi: 10.3109/17483107.2014.981874] [Medline: 25391221]
45. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, American Heart Association Stroke Council, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Quality of Care and Outcomes Research. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2016 Jun;47(6):e98-169. [doi: 10.1161/STR.0000000000000098] [Medline: 27145936]

Abbreviations

ADL: activities of daily living
FMA-UE: Fugl-Meyer Assessment for upper extremity
GRASP: Graded Repetitive Arm Supplementary Program
MAL: Motor Activity Log
MCID: minimal clinically important difference
REDCap: Research Electronic Data Capture
SIS: Stroke Impact Scale
UE: upper extremity
VR: virtual reality

Edited by N Zary; submitted 24.02.22; peer-reviewed by R Proffitt, M Naeemabadi, JM Pastora Bernal; comments to author 07.04.22; revised version received 27.05.22; accepted 21.07.22; published 27.09.22.

Please cite as:
Hernandez A, Bubyr L, Archambault PS, Higgins J, Levin MF, Kairy D
Virtual Reality–Based Rehabilitation as a Feasible and Engaging Tool for the Management of Chronic Poststroke Upper-Extremity Function Recovery: Randomized Controlled Trial
JMIR Serious Games 2022;10(3):e37506
URL: https://games.jmir.org/2022/3/e37506
doi:10.2196/37506
PMID:36166289

©Alejandro Hernandez, Liudmila Bubyr, Philippe S Archambault, Johanne Higgins, Mindy F Levin, Dahlia Kairy. Originally published in JMIR Serious Games (https://games.jmir.org), 27.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Speech Processing as a Far-Transfer Gauge of Serious Games for Cognitive Training in Aging: Randomized Controlled Trial of Web-Based Effectivate Training

Gal Nitsan[1,2*], MA; Shai Baharav[2*], BA; Dalith Tal-Shir[2], BA; Vered Shakuf[3], PhD; Boaz M Ben-David[2,4,5], PhD

[1]Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel
[2]Baruch Ivcher School of Psychology, Reichman University (IDC), Herzliya, Israel
[3]Department of Communications Disorders, Achva Academic College, Arugot, Israel
[4]Toronto Rehabilitation Institute, University Health Networks, Toronto, ON, Canada
[5]Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Boaz M Ben-David, PhD
Baruch Ivcher School of Psychology
Reichman University (IDC)
P.O. Box 167
Herzliya, 4610101
Israel
Phone: 972 584004055
Email: boaz.ben.david@idc.ac.il

Abstract

Background: The number of serious games for cognitive training in aging (SGCTAs) is proliferating in the market and attempting to combat one of the most feared aspects of aging—cognitive decline. However, the efficacy of many SGCTAs is still questionable. Even the measures used to validate SGCTAs are up for debate, with most studies using cognitive measures that gauge improvement in trained tasks, also known as *near transfer*. This study takes a different approach, testing the efficacy of the SGCTA—*Effectivate*—in generating tangible *far-transfer* improvements in a nontrained task—the Eye tracking of Word Identification in Noise Under Memory Increased Load (E-WINDMIL)—which tests speech processing in adverse conditions.

Objective: This study aimed to validate the use of a real-time measure of speech processing as a gauge of the far-transfer efficacy of an SGCTA designed to train executive functions.

Methods: In a randomized controlled trial that included 40 participants, we tested 20 (50%) older adults before and after self-administering the SGCTA *Effectivate* training and compared their performance with that of the control group of 20 (50%) older adults. The E-WINDMIL eye-tracking task was administered to all participants by blinded experimenters in 2 sessions separated by 2 to 8 weeks.

Results: Specifically, we tested the change between sessions in the efficiency of segregating the spoken target word from its sound-sharing alternative, as the word unfolds in time. We found that training with the SGCTA *Effectivate* improved both early and late speech processing in adverse conditions, with higher discrimination scores in the training group than in the control group (early processing: $F_{1,38}$=7.371; P=.01; η_p^2=0.162 and late processing: $F_{1,38}$=9.003; P=.005; η_p^2=0.192).

Conclusions: This study found the E-WINDMIL measure of speech processing to be a valid gauge for the far-transfer effects of executive function training. As the SGCTA *Effectivate* does not train any auditory task or language processing, our results provide preliminary support for the ability of *Effectivate* to create a generalized cognitive improvement. Given the crucial role of speech processing in healthy and successful aging, we encourage researchers and developers to use speech processing measures, the E-WINDMIL in particular, to gauge the efficacy of SGCTAs. We advocate for increased industry-wide adoption of far-transfer metrics to gauge SGCTAs.

(JMIR Serious Games 2022;10(3):e32297) doi:10.2196/32297

XSL•FO
RenderX

KEYWORDS

aging; cognitive aging; cognitive games; serious games; speech processing; spoken language processing; eye tracking; visual world paradigm

Introduction

Background

The age distribution of the world's population is projected to dramatically shift over the next few decades as improved health care continues to extend life expectancy [1]. By 2050, more than one-fourth (27%) of the European population is expected to be aged >65 years. Although medicine can prolong relative physical health [2], offsetting age-related changes in cognitive health is growing in importance [3]. Although recent literature suggests that cognitive measures may inflate age-related decreases in performance [4-6] and disregard an increase in crystallized intelligence (eg, general knowledge and vocabulary [7]), a decrease in cognitive performance is one of the most feared aspects of aging [8]. Consequently, there is growing pressure to prolong active and healthy aging—*successful aging* [9]—by targeting age-related changes in cognitive abilities such as memory and executive functions (EFs; eg, inhibition and working memory [WM]) [10,11].

Numerous serious games for cognitive training in aging (SGCTAs) are being developed to mediate age-related cognitive changes. However, there is much debate in the literature regarding their efficacy. In fact, in 2014, a total of 2 teams of researchers published contradictory open letters. The first letter by a group of 70 scientists refuted the efficacy of such training [12]. The second letter by another group of 133 scientists claimed the opposite, supporting the benefit of cognitive training [13]. These letters were followed by an extensive review [14] suggesting that SGCTAs, in general, can improve performance on the trained game and associated activities—*near transfer*. However, the review cautions that there is insufficient evidence to suggest that these changes can *generalize* to activities that are not directly associated with the game—*far transfer*. As cognitive training games and interventions are a "means to enhance performance on other tasks" [14], it seems critical to measure their effects using far-transfer measures that gauge daily activities. Such far-transfer measures would gauge cognitive abilities through performance on a different task that is mediated by the trained cognitive functions.

Effective communication and speech perception play an extensive role in many daily activities and have evident effects on general health and well-being [15]. Difficulty in understanding speech in adverse conditions (eg, noisy background or while conducting another task) forms one of the most prevalent complaints among older adults [16]. These difficulties decrease the participation of older adults in social and professional interactions, thus limiting their independence and increasing feelings of loneliness. Growing evidence suggests that a decrease in speech processing, in turn, has a negative effect on mental health, general well-being, and even longevity [17-21]. The social restrictions imposed by the COVID-19 pandemic further limit interactions and other opportunities for cognitive exercise (eg, work and volunteering). Indeed, current restrictions have been found to increase loneliness and depression in older age [22], even after vaccinations were made available [23] and severe social restrictions were lifted [24]. These, together with limited access to health care services as a result of the pandemic [25,26], illustrate the necessity to create effective SGCTAs that can directly affect spoken communication in adverse conditions, even while social distancing. We suggest that testing speech processing as a far-transfer task could demonstrate the impact of training with SGCTAs, on the daily lives of older adults.

In this exploratory study, we used an eye-tracking paradigm to assess whether training EFs with the SGCTA *Effectivate* generalizes to improved speech processing in adverse listening conditions for older adults. This will serve to validate a real-time measure of speech perception—Eye tracking of Word Identification in Noise Under Memory Increased Load (E-WINDMIL)—as a gauge of far-transfer efficacy of SGCTAs designed to train EFs and provide a case study for developers and academics on the use of far-transfer metrics.

EFs and Speech Processing

Many SGCTA developers are targeting EFs because of their prominent role in healthy cognitive aging (refer to the seminal work by Salthouse [27,28]). EFs, which include WM and inhibition, enable active maintenance and manipulation of bottom-up information with top-down information in memory, especially during the performance of a concurrent task [29-31]. The literature suggests that individuals with better EFs are able to hold more incoming information and incorporate and manipulate it more easily, even under adverse conditions such as distractions (ie, external noise) and memory preload (ie, remembering the context of a conversation [32,33]). Therefore, it is not surprising that EFs play a significant role in speech processing [34].

Consider a scenario in which an older adult is driving his grandson in a car and radio music is playing. The grandson says, "grandpa, have you seen the DOLL?" The older listener must perform the following tasks:

1. Segregate the spoken message from the background radio noise (task-irrelevant) stream as it unfolds in time
2. Inhibit the activation of competing (similar-sounding) words in the mental lexicon (eg, */DO/* sounds in words such as *DOG*) while increasing the activation of the word DOLL, as the sound */L/* unfolds in time
3. Allocate enough resources for the activities mentioned previously from a limited cognitive resource pool that is already depleted by the concurrent task of driving

As mentioned previously, EFs, especially WM, are essential to perform this complex task and have been shown to be affected by aging in the following ways: (1) stream segregation slows with aging, (2) decrease in the efficiency of inhibition impairs the ability to reject incorrect lexical candidates, (3) decrease in cognitive resources can impair speech perception, and (4)

XSL•FO
RenderX

age-related hearing loss distorts the perception of bottom-up signals.

First, stream segregation slows with aging [35]. Reduced WM capacity has been linked to limitations in inhibition [36]. This affects the ability to separate relevant speech from irrelevant background noise. For example, in a study by Janse [37], when speech was presented in background noise, poor inhibitory abilities led to greater interference by the competing noise, which impaired speech perception in older adults.

Second, an age-related decrease in the efficiency of inhibition [15,38] impairs the ability to reject incorrect lexical candidates as the context unfolds in time [39,40].

Third, age-related decreases in cognitive resources, specifically in EFs and WM [38], can impair speech perception, as suggested by the *Framework for Understanding Effortful Listening* [41]. The *Framework for Understanding Effortful Listening* is an adaptation of the capacity model of attention by Nobel laureate Daniel Kahneman, which conceptualizes the relationship between mental resource capacity and cognitive demands. According to this model, mental resources have limited capacity. The presence of background noise or another resource-consuming task (eg, driving) can impede and slow down speech processing for people with lower resource capacities.

Finally, age-related hearing loss distorts the perception of bottom-up signals, providing impoverished input to the central nervous system. To mitigate these effects, older adults rely heavily on the linguistic context in word recognition, often to an even greater degree than younger adults [34,42,43]. Efficient context processing depends on WM capacity and information processing speed [44]. As mentioned previously, these capabilities decline with age [28,45] and can affect older adults' ability to use context during word recognition. Depleted WM capacity can also affect the ability to temporarily remember words from a given linguistic context for later use [46].

In summary, cognitive performance is intertwined with speech perception, especially in older age. Age-related difficulties in speech perception are not only affected by reduced cognitive abilities but can also accelerate the rate of cognitive decline. A total of 2 Lancet reports [15,47] on dementia prevention highlighted improving auditory and speech accessibility as the number one modifiable risk factor in middle to late life. In fact, the relative weight of speech accessibility in preventing dementia is estimated to be higher than in tackling smoking, diabetes, hypertension, and obesity altogether. As Lin [48] suggests in his *Aging and Cognitive Health Evaluation in Eldersmodel* [49], degraded speech processing affects cognitive resilience in aging by decreasing physical activities, social interactions, communication, and related brain functions. Hence, it is plausible to assume that training EFs should enable participants to juggle informational weight more gracefully and process speech faster, in turn, improving their quality of life and well-being.

Speech as a Far-Transfer Measure of Cognitive Training Using Eye Tracking

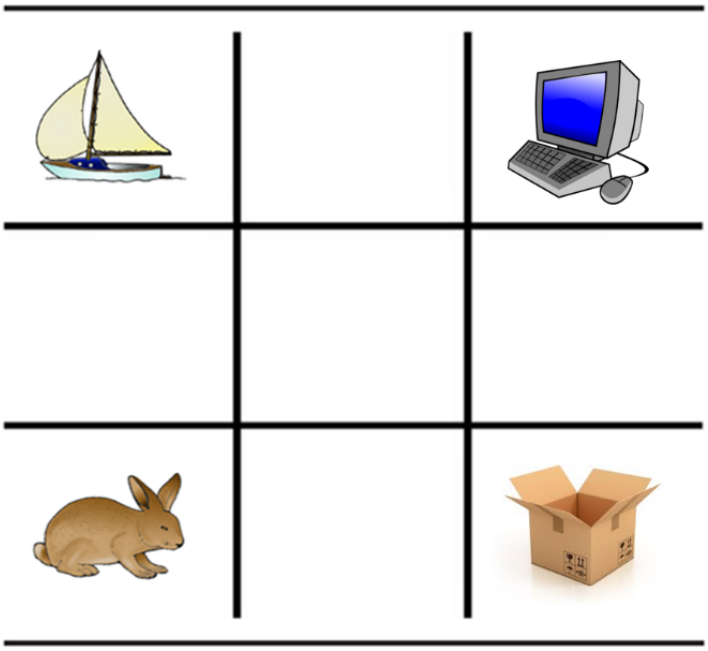
To test the effect of cognitive training on speech processing, this study used eye tracking. We used a noninvasive infrared light source and high-precision camera that collects reflections from the eye and records the exact location of the eye gaze on the display at a rate of 500 samples per second. As the word unfolds in time, eye gaze data are time locked with what is being heard by the listener. By recording the participant's eye movements in relation to the visual display and auditory stimuli, eye tracking provides a highly sensitive and continuous measure of spoken word processing. Unlike overt non–real-time responses (participant verbally or physically responding *after* the word has been heard), the covert rapidity of an eye movement allows one to determine the point in time at which the listener is able to isolate the target word from its competitors through the difference in fixations on the target and competitor over time. Although non–real-time responses, such as pointing at the screen, may be affected by age-related motor slowing, covert eye movement speed and accuracy are relatively unaffected [50].

To specifically gauge the cognitive mechanisms involved in speech processing under adverse conditions, our laboratory adapted the *Visual World* eye-tracking paradigm [51] to include a concurrent task (increasing memory load) and noise (increasing distractions), creating the E-WINDMIL [44]. In E-WINDMIL, listeners hear Hebrew sentences such as "point at the box" while viewing a visual display on a computer screen that contains 4 objects. In this example (Figure 1), the display shows a picture of the named object heard by the participant, *box /ar.gaz/*, along with three other objects: a phonological competitor (eg, an onset competitor that shares the first syllable with the target, *rabbit /ar.nav/*) and 2 additional objects that are neither semantically nor phonologically related to the heard target object or its name. Participants are asked to touch the picture of the object as quickly and accurately as possible while their eye gaze is recorded. Rather than analyzing the slower overt touch response, only the eye gaze is taken into account in later analysis.

As real-life speech processing is often accompanied by other tasks, before the onset of the spoken instructions, participants are also asked to retain in memory either 1 or 4 digits (low or high memory load, respectively) for later recall. A discrimination score, which is the difference between the proportion of eye gaze fixations to the target (image representing the heard word) and the phonological competitor, was used to assess the 2 groups. The higher the difference, the more efficient the listener is in discriminating the spoken target from its signal-sharing competitor. Using the same eye-tracking paradigm, Nitsan et al [52,53] showed that listeners with larger WM capacity were able to identify the target word (and reject the signal-sharing competitor) earlier than a matched group with lower capacity. These findings suggest that improving one's cognitive capacity might improve speech processing in adverse conditions.

XSL•FO
RenderX

Figure 1. An example of the experimental display. The target word in this example, /ar.nav/ (rabbit), is represented in the bottom-left corner. The phonological competitor, /ar.gaz/ (box), is represented in the bottom-right corner. The words /si.ra/ and /max.ʃev/ (boat and computer, respectively) are unrelated distractor.



This Study

A total of 2 groups of older adults were tested *twice* on the E-WINDMIL speech processing task. One group received no cognitive training, whereas the other group followed the *Effectivate* SGCTA training protocol for 6 weeks. We aimed to test whether a short training period using *Effectivate* would engender a significant far-transfer change in speech processing ability. As the tested SGCTA does not involve any type of auditory training, improved performance on the E-WINDMIL speech processing task would provide strong support for the far-transfer effect and demonstrate the use of far-transfer measures in gauging training success for the validation of SGCTAs.

We hypothesized that if the tested SGCTA, *Effectivate*, improves generalized EFs, speech processing in adverse conditions, as measured by performance on the E-WINDMIL, would improve for the training group. Specifically, the difference in discrimination scores between the training group and control group would not be significant in the first test session, although an advantage for the training group would be found in the second test session (after training). This would suggest that training had a significant impact on real-time speech processing, above and beyond practice with the E-WINDMIL task.

Methods

Participants

A total of 54 older adults were recruited via phone calls from the Reichman University's older adult research volunteer group and randomly assigned to either the cognitive training or control group. Although the groups cannot be said to reflect the diversity of the global older adult population, they are representative of the population residing in central Israel, where the study was conducted. Of these 54 individuals, 8 (15%) did not return for the second eye-tracking session, and 6 (11%) were excluded because of failure in eye movement recording or loss of eye-tracking signal. Recruitment was continuous for the duration of 6 months. Owing to the COVID-19 pandemic, participant recruitment and data collection were limited and terminated earlier than expected. The training group comprised 50% (20/40) of older adults (mean age 65.65, SD 4.8 years; 14/20, 70% were women). The control group comprised 50% (20/40) of older adults (mean age 69.05, SD 3.8 years; 13/20, 65% were women) from the study by Baharav et al [54]. All participants met the research inclusion criteria (refer to Textbox 1 for details). As shown in Table 1, the 2 groups had similar gender distribution (P=.74). Hearing acuity (pure tone average), years of education, and forward digit span scores did not differ significantly between the 2 groups (P=.51, P=.74, and P=.76, respectively). However, participants in the training group were slightly younger (t_{38}=2.48; P=.02). All the participants provided written informed consent.

XSL•FO
RenderX

Textbox 1. Inclusion criteria for participant recruitment.

Language background

High proficiency Hebrew speakers (no early bilinguals were included), assessed by a self-report and a score within the normal range in the Wechsler Adult Intelligence Scale–3 Hebrew vocabulary subtest

Hearing

Symmetrical air conduction hearing thresholds, expressed as pure tone averages of ≤25 dB hearing level in each ear (0.5 kHz, 1 kHz, and 2 kHz), and no reported history of auditory pathology

Vision

Normal or corrected to normal visual acuity and color vision, assessed by the Landolt C charts and the Ishihara charts

Cognition: working memory

Clinically normal scores for their age range on the Montreal Cognitive Assessment cognitive screening test and on the forward and backward digit span subtests (Hebrew version of Wechsler Adult Intelligence Scale–3 [46])

Table 1. Demographic characteristics (N=40).

Characteristics	Training group (n=20)	Control group (n=20)	Group comparison		
			t test[a] (*df*)	Chi-square (*df*)	*P* value
Age (years), mean (SD)	65.65 (4.848)	69.041 (3.605)	2.478 (38)	N/A[b]	.02
Gender (women), n (%)	14 (70)	13 (65)	N/A	0.4 (1)	.74
Hearing (across 0.5 kHz, 1 kHz, and 2 kHz), mean (SD)	16.79 (4.939)	17.85 (4.913)	0.672 (37)	N/A	.51
Education (years), mean (SD)	16.42 (2.244)	16.18 (2.69)	0.339 (34)	N/A	.74
Digit span, mean (SD)	9.9 (1.714)	9.75 (1.333)	0.309 (38)	N/A	.76

[a]The *t* test was 2-tailed.

[b]N/A: not applicable.

Ethics Approval

Ethics approval for this study was obtained from the Reichman University (Interdisciplinary Center Herzliya) institutional review board (P_1920119). This study was conducted in line with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1).

Stimuli

Auditory Stimuli

Auditory stimuli were taken from the study by Nitsan et al [52,53] and contained both the object names describing the visual stimuli and the sentence, “point at the ___ [target word]” in Hebrew using a plural non–gender-specific form. All object names were disyllabic. The average target word duration, including the Hebrew definition article *ha-* (the), was 1078 (SD 91) milliseconds. The root mean square intensity was equated across all recorded sentences. Files were mixed with a continuous speech spectrum noise at a fixed 0 dB signal-to-noise ratio based on values for the discrimination timeline in the study by Ben-David et al [55]. Stimuli were presented binaurally at 50 dB above the individual pure tone average via a MAICO MA-51 (MAICO) audiometer using TDH 39 supra-aural headphones (Telephonics).

Visual Display

In each trial, the participants were presented with a 3×3 grid with 4 images of objects positioned at the grid corners (Figure 1). The stimuli (images) were previously used in the studies by Hadar et al [56] and Nitsan et al [52,53] and were confirmed to be clearly identifiable and highly familiar. In all the trials, 1 of the 4 image names corresponded with the spoken target word. In critical trials, a second image name was a phonological competitor: sharing the initial syllable (onset overlap) or the final syllable (offset overlap) with the spoken target word. The remaining 2 objects presented on the screen were phonologically and semantically unrelated to both the target word and phonological competitor. In addition to critical trials, filler trials were used to diminish participant expectation of onset phonetic resemblance between the depicted object names. In filler trials, all 3 distractors were phonologically and semantically unrelated to the target word.

The original image database was divided into two to create 2 image sets, which were counterbalanced between participants for testing sessions 1 and 2. Within each testing session, objects were presented twice: once as a critical trial and once as a filler trial in which one of the 2 phonetically *unrelated* items was used as the target word. To prevent implicit spatial learning within a single testing session, object positions on the screen were rotated at each presentation.

Procedure

Overview

The study comprised 2 experimental sessions, all conducted individually in a dedicated experimental laboratory complex at the Reichman University. In the first session, participants signed

XSL•FO
RenderX

an informed consent form, and the inclusion criteria measures were collected. The E-WINDMIL paradigm was administered (as presented in the following section) to determine the participants' baseline performance. To maintain experimenter blindness to the conditions, 2 different research assistants conducted the experiment. One research assistant conducted the E-WINDMIL and auditory testing, and the other research assistant assigned participants to each group and presented the participants in the training group with a web address providing access to the *Effectivate* SGCTA and instructed them to train at least three times a week for a duration of 5 weeks, after which they returned for the second experimental session. They were called once a week to verify *Effectivate* training. In the control group, participants were asked to maintain their daily routine and return within 2 to 4 weeks. In the second session, the same E-WINDMIL task was administered and the participants were debriefed. All participants were aware of the academic affiliation of the researchers. Participants in the experimental condition were not blinded to the name of the SGCTA company; however, the product was still in the beta stages and, as such, was not publicly available or marketed at the time.

E-WINDMIL Paradigm

The experiment was administered individually in a dedicated sound-attenuated booth (IAC Acoustics). Participants were seated 60 cm away from the computer screen, with their heads placed on the designated eye-tracker chin rest to minimize head movement. Each participant's dominant eye was calibrated to ensure that their real-time eye gaze position was recorded throughout the course of the trial. A table-mounted SR Eyelink 1000 eye tracker (SR Research Ltd) in the *tower mount* configuration was used. Eye gaze position was recorded using the Eyelink software at a rate of 500 Hz.

Trials began with a visual cue of a black *play* triangle centered on the screen, immediately followed by the auditory presentation of either 1-digit preload (low-load condition) or 4-digit preload (high-load condition) through headphones. Participants were told to memorize these digits (in the order presented) for later recall. Subsequently, a 3×3 grid with the 4 images appeared (Figure 1). Participants were given 2 seconds to view the object positions, after which a fixation cross appeared in the center of the screen. Once the participants pressed the fixation cross to initiate the trial, the instruction sentence, "point at the ___ [target word]," was presented via the headphones. Selection of a named object was indicated by touching the object's picture on the touch screen. Following the participant's selection of a stimulus, a visual feedback signal appeared in the square of the selected image: red highlight for an incorrect answer or green highlight for a correct answer. Finally, the visual display was cleared, and a visual cue of a black circle appeared on the screen, signaling participants to recall aloud the digit preload from the beginning of the trial. Then, the experimenter coded the response (either correct or incorrect) in real time. Participants were instructed that the speed and accuracy of both the object selection and digit recall were equally important.

In a given testing session, participants completed 68 trials, split into 2 trial blocks for each digit preload condition (low load: 1 digit; high load: 4 digits). Each condition contained 34 trials, of which 2 (6%) were practice trials, and 32 (94%) were experimental trials. The 32 trials in each condition were split such that 16 (50%) were *filler* trials, indicating that the target object's name did not share any phonology with the surrounding objects, and 16 (50%) were *critical* trials, indicating that the target object's name shared phonology with a surrounding object name. 50% (8/16) were phonological onset competitors (eg, */ar.nav/-/ar.gaz/*), and 50% (8/16) were phonological offset competitors (eg, /xalon/-/balon/).

Although participants in the experimental group were aware of the intervention, the *Visual World* covert eye-tracking design was found to account for participants trying to outperform in an overt choice of the target (eg, with a button press). In other words, participants cannot control eye gaze fixations toward the alternatives versus fixations toward the target once saccades have been initiated. Indeed, in the visual world paradigm, eye movements were affected by implicit task goals and relatively immune to intentions and social desirability [57-59].

SGCTA Effectivate

Following baseline testing, participants in the training group completed at-home web-based training, using a PC or tablet. A minimum of 15 training sessions were completed with approximately 8 minutes of active training per session (range 3-15 minutes). Each training session comprised 2 to 10 exercises, which were selected from a bank of 10 tasks. The difficulty level was individually adjusted for each participant and calibrated separately for each task using various measures (eg, exposure time, reaction time window, and number of objects). Each training task targeted at least one of the following cognitive functions: processing speed, WM, executive control, attentional control, sustained attention, spatial attention, binding, semantic memory, and training of several mnemonic methods. Figure 2 presents an example of such an exercise.

XSL•FO
RenderX

Figure 2. An example of a slide from the Effectivate serious game for cognitive training in aging—the exercise, The Last One Counts, is based on the 'n-back' task. In this exercise, the users were presented with a sequence of shapes and asked to decide whether each shape is identical to the one previously presented. Task difficulty changed gradually by updating different parameters, such as exposure latencies. In advanced levels, users were asked to decide whether the current shape is the same as, different from, or partially similar to the previously presented one. This additional level of complexity requires users to segregate the item's different features (ie, color and shape) to selectively focus on some and inhibit others.



Results

Response Accuracy

Table 2 presents the accuracy percentage for each experimental condition—the percentage of trials in which participants both correctly selected the corresponding object on the visual display (indicating correct spoken word recognition) and correctly recalled the preload digits (indicating correct digit recall). A Mann-Whitney independent-sample nonparametric test confirmed that WM load, test session, and participant group did not have significant effects on accuracy, with $P>.17$ for all 4 tests.

Table 2. Mean percentage (and SDs) of trials in which the target word was correctly selected and digits were correctly recalled[a].

Participant group and WM[b] load	First session (%), mean (SD)	Second session (%), mean (SD)
Training		
Low	99.4 (2.8)	98.1 (4.6)
High	91.2 (14.1)	88.1 (15.4)
Control		
Low	98.9 (3.7)	99.4 (4.6)
High	87.5 (13.4)	92.6 (10.7)

[a]Low working memory and high working memory indicate the two preload conditions, 1 digit and 4 digits, respectively.

[b]WM: working memory.

Eye Gaze Analysis

We analyzed target discrimination scores (following the methodology of previous studies [60-63]) reflecting the listeners' ability to discriminate the target word from its phonological competitor. The proportion of fixations on the competitor was subtracted from the proportion of fixations on the target within 250-millisecond time bins, starting from 250 milliseconds after the word onset to 1500 milliseconds. In this measure, the higher the value, the better listeners can discriminate the target from its competitor; values approaching 0 reflect an inability to discriminate between the target and competitor objects. Mixed-design repeated-measures ANOVAs were conducted for each 250-millisecond time bin, with three within-participants factors—WM load (high and low), test session (first and second), and condition (onset vs offset sound sharing)—and one between-participant factor—participant group (training and control). In each analysis, planned comparisons compared the effect of the participant group on discrimination scores in the first and second test sessions to verify whether differences between groups were related to the intervention (ie, significant effect only in the second session). Significant interactions of the test session with the participant group were noted in two of the five tested time bins: early processing 250 to 500 milliseconds and late processing 1250 to 1500 milliseconds, as discussed in the following section (Figures 3 and 4). The remaining three time bins (500-750 milliseconds, 750-1000 milliseconds, and 1000-1250 milliseconds) did not show any significant interaction; thus, they will not be discussed further.

XSL•FO
RenderX

Figure 3. First test session. Mean target discrimination scores (with SE bars) for the training and control groups. Target discrimination scores are the proportion of fixations on the competitor subtracted from the proportion of fixations on the target within 250-millisecond time bins, starting from 250 milliseconds after the word onset to 1500 milliseconds.

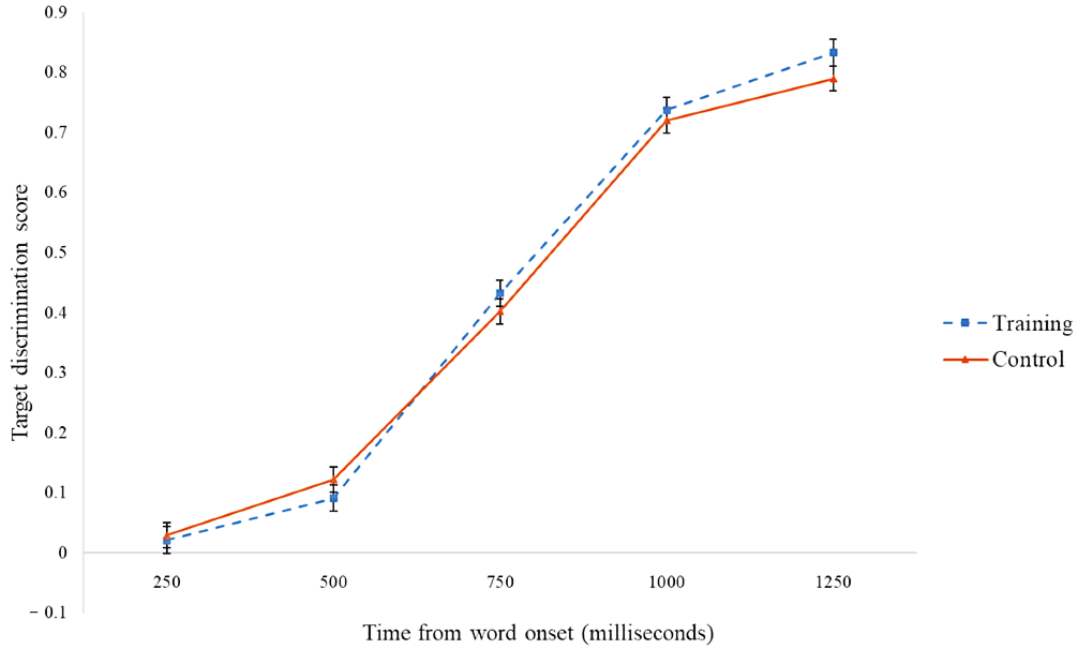
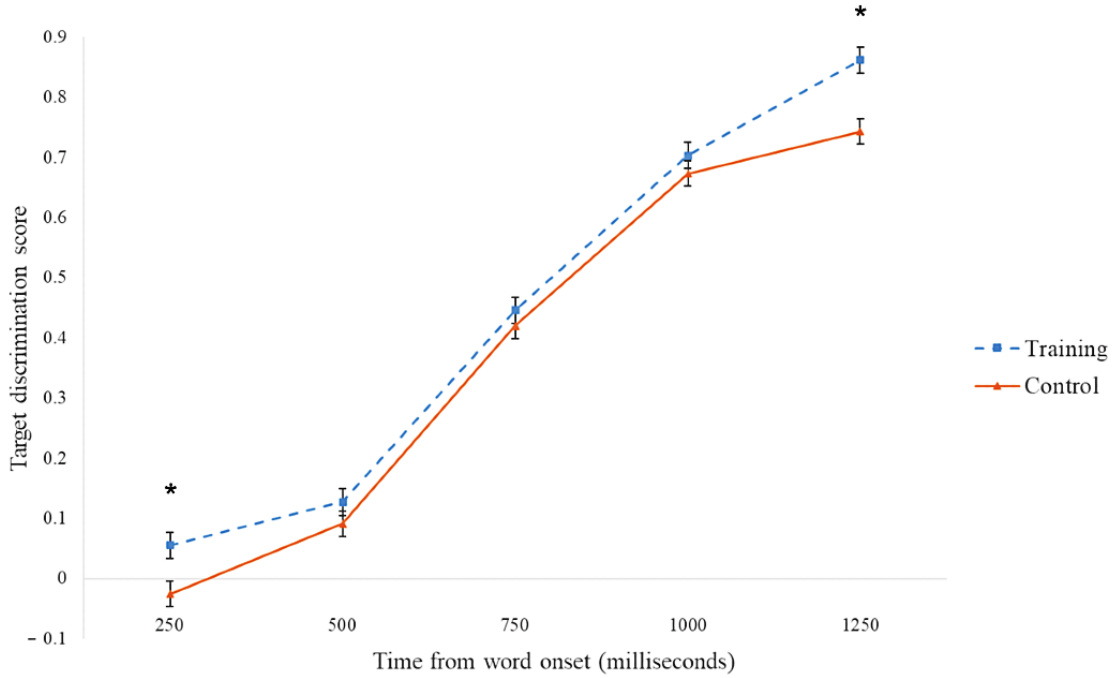


Figure 4. Second test session. Mean target discrimination scores (with SE bars) for the training and control groups. Target discrimination scores are the proportion of fixations on the competitor subtracted from the proportion of fixations on the target within 250-millisecond time bins, starting from 250 milliseconds after the word onset to 1500 milliseconds. *Significant effect.



Early Processing: 250- to 500-Millisecond Time Bin

The interaction between the test session and the participant group was found to be approaching significance ($F_{1,38}$=3.881; P=.06; η_p^2=0.093). Planned comparison indicated no significant difference between the 2 groups in the first session ($F_{1,38}$=0.056; P=.81), whereas the second session produced higher discrimination scores in the training group ($F_{1,38}$=7.371; P=.01; η_p^2=0.162). This suggests that improved performance can be related to the intervention itself. The effect of the participant group was marginally significant ($F_{1,38}$=3.048; P=.07; η_p^2=0.085), with slightly higher discrimination scores in the training group (as in the previous analysis, this difference can be related to the second session) and no significant main effect for test session ($F_{1,38}$=0.224; P=.64). No significant triple interactions were found for the participant group, test session, or any other tested variables (WM load or condition).

XSL•FO
RenderX

Late Processing: 1250- to 1500-Millisecond Time Bin

A significant interaction between the test session and the participant group was found ($F_{1,38}$=4.220; P<.05; η_p^2=0.100). Planned comparisons indicated that although the 2 groups did not significantly differ in the first session ($F_{1,38}$=1.689; P=.20), the second session showed higher discrimination scores in the training group than in the control group ($F_{1,38}$=9.003; P=.005; η_p^2=0.192), suggesting that improved performance could be related to the intervention itself. A significant main effect of the participant group was noted ($F_{1,38}$=6.722; P=.01; η_p^2=0.150), with higher discrimination scores in the training group (emanating from higher scores in the second session) and no significant main effect for the test session ($F_{1,38}$=0.108; P=.74). No significant triple interactions were found for the participant group, test session, or any other tested variables (WM load or condition).

In summary, in early and late processing (250-500 milliseconds and 1250-1500 milliseconds after word onset, respectively), performance did not differ between the 2 groups in the first test session. However, in the second test session after training with the *Effectivate* SGCTA, the training group surpassed the control group. These effects can be taken to suggest that the SGCTA training improved performance, over and above any effect of test-retest repetition.

Discussion

Principal Findings

In this exploratory study, we aimed to validate an eye-tracking paradigm, the E-WINDMIL, which tests real-time speech processing in adverse conditions as a gauge for the far-transfer efficacy of SGCTAs. Specifically, we tested whether training EFs in the visual modality with the SGCTA *Effectivate* generalizes to improved speech processing in adverse listening conditions (auditory modality) for older adults.

The training group, with 50% (20/40) of the older adults, was tested before and after 6 weeks of training on the SGCTA *Effectivate*. The control group, with another 50% (20/40) of the older adults, did not undergo any specific cognitive training. Before training, no significant differences in E-WINDMIL performance were noted between the control group and the training group. However, after training with *Effectivate*, the training group outperformed the control group in early word processing (indicated by eye movements, 250-500 milliseconds after word onset) and late word processing (1250-1500 milliseconds after word onset). The early processing advantage may suggest improved stream segregation between the spoken target word and noise [52,53], when WM was otherwise occupied. The late processing advantage alludes to improved decision-making processes (ie, using accumulated evidence) once the word had been completely heard [55]. Our results provide early support for the efficacy of the E-WINDMIL speech processing paradigm as a far-transfer measure of cognitive training with the SGCTA *Effectivate*. This is of special interest as the tested SGCTA did not train any auditory task or spoken language processing.

Speech Processing as a Far-Transfer Gauge for Cognitive Training

Challenges in determining the effectiveness of any cognitive intervention stem from the ongoing debate: Do we use near-transfer or far-transfer metrics? [14] In other words, is it sufficient to indicate improved performance on the trained task or should research indicate improved performance on a *daily* task, far from training, to suggest the *generalizability* of training? This exploratory study demonstrates the efficacy of using a far-transfer measure that involves speech processing in adverse conditions to discern the impact EF training has on daily life activities.

Speech processing in adverse conditions presents an excellent gauge of the generalizability of cognitive training. As speech processing is resource demanding, the fewer resources listeners have, the more they will be affected by adverse conditions such as background noise. Speech processing involves holding ongoing speech strings in memory and integrating words and phrases to create coherent meaning; thus, it is considered to be dependent on WM and other attentional resources [52,56]. The results of this study suggest the prowess of training to create a generalized cognitive effect, as a few weeks of training on the *Effectivate* SGCTA was sufficient to improve speech processing in adverse conditions (above and beyond test-retest learning effects).

This improvement can be interpreted in light of the crucial role of EFs, especially WM, in speech processing in adverse conditions. According to the Ease of Language Understanding model [44], explicit WM resources are drawn from a central pool to compensate for the loss of automatic matching between the input and lexical representations when the sound input is degraded by adverse listening conditions. Other studies have demonstrated a direct link between WM capacity and the ability to inhibit irrelevant information. This ability is necessary to separate the speech signal from background noise and reject competing words in the mental lexicon. Thus, our results suggest that training EFs using SGCTAs might have a generalized effect on real-life daily tasks. Returning to our example in the introduction, with improved WM capacity, the older adult will be better able to understand his grandson saying, “Grandpa, have you seen the DOLL?” rather than *DOG* (sound-sharing alternative) while driving a car (WM load) with the radio playing (adverse listening conditions) in the following ways:

1. Improve speech segregation—separating the spoken message from the background task-irrelevant noise (eg, radio and engine noise)
2. Effectively inhibit the activation of competing similar-sounding words in the mental lexicon (eg, *DOG*)
3. Allocate enough resources to use context and information in long and short memory from a cognitive resource pool, which is already depleted by the concurrent task of driving

Given the pivotal role of speech processing in successful aging [64], this change may have a lasting positive impact on the quality of life in older age.

XSL•FO
RenderX

E-WINDMIL as a Far-Transfer Gauge for Cognitive Training

The advantage of using the adapted visual world paradigm E-WINDMIL lies in its increased ecological validity—measuring a daily task (speech processing) that is important to the perseverance of well-being and performance in older age [17-19]. Eye tracking used by E-WINDMIL is better suited to test older adults' speech processing than more traditional speech tests involving overt responses, such as verbal or keypress response. It is not influenced by an age-related slowing of motor speed, which often affects non–real-time speech tests [65]. Unlike many other speech processing tasks that assess the processing of a single word in *ideal listening* conditions, the E-WINDMIL asks listeners to retain digits for later recall (a task designed to weigh on WM resources) while presenting speech in noise. In this way, the E-WINDMIL paradigm acknowledges that speech in real-life scenarios is often experienced along with noise while the listener is engaged in other cognitively demanding tasks (eg, following the context of the sentence as it unfolds and driving). Moreover, eye tracking has been shown to be a sensitive measure for speech processing in various studies, suggesting that speech processing is costly in terms of WM processing and perhaps even mediated by it [29,34,56,66,67].

Training-Related Advantage in the 1250- to 1500-Millisecond Time Bin

Previous adaptations of E-WINDMIL found eye tracking to be very sensitive to differences in cognitive reserve. Hadar et al [56] found that minimizing available cognitive resources can slow down processing in this task. Nitsan et al [52,53] found that individuals with higher cognitive reserve outperform individuals with lower reserve while using E-WINDMIL. This advantage, attributed by the authors to the use of cognitive resources for speech processing, was indicated in the later time bins—similar to the current findings. Other studies also found that differences in cognition were indicated in later word processing with older listeners in particular [61,68]. A recent study by Harel-Arbeli et al [46] attributed the advantages seen in later time bins to decision-making processes. In their study, using a similar eye-tracking paradigm, the spoken target word was preceded by a spoken predictive context presented in a quiet environment. An advantage of young adults over older adults, based on the age-related difference in cognitive resources, was present mainly in the late time bin when the full word had been spoken. Taken together, it appears that improved processing in the late time bin may reflect improved cognitive resources (eg, WM and inhibition).

Training-Related Advantage in the 250- to 500-Millisecond Time Bin

The current data also indicated training-related advantages in processing during early time bins, when only the first phoneme of the word is being processed. This suggests that cognitive training improved target word stream formation and auditory stream segregation between the target word and noise [55,69]. Indeed, this early process of stream segregation has been linked not only to sensory processes but also to the deployment of cognitive resources. Cognition is necessary for the inhibition of the noise stream and selective focus on the target word stream, leading to stream segregation [70,71]. Stream segregation is essential for speech processing and represents one of the major hurdles for older adults in social interactions [72]. Indeed, age-related auditory sensory degradation can specifically impair processes related to stream segregation in aging [73]. This early time bin training advantage may also be related to the early benefits noted in the literature as a result of removing background noise [60] and increasing the lexical frequency of the spoken word [74] using similar eye-tracking paradigms. In summary, the performance advantage in the early time bin associated with SGCTA training may reflect an increase in cognitive reserve.

Caveats and Future Studies

This study should be taken as a first step in supporting the effectiveness of the tested SGCTA, and it does not serve as a recommendation or suggestion to use SGCTAs in general or specifically the *Effectivate* SGCTA. This study was ongoing at the beginning of the COVID-19 pandemic and was halted because of national quarantine. Therefore, we were unable to amass a larger group of participants. Moreover, we were unable to recruit an active control group to undergo an alternative form of cognitive training. Future studies should attempt to replicate the results with an active control to ensure that the observed effects were not related to possible social desirability or lack of participant blinding but to the specific cognitive training, *Effectivate*. However, we note that the experimenters administering the study were blinded to the condition, and the experimental tool was relatively immune to social desirability. Such replications should also more carefully match participants across all groups. Indeed, on average, participants in the control group were older by a few years than those in the training group. We also note that participants in this study did not form a representative sample of the older adult population, specifically given the cognitive and linguistic inclusion criteria. Although future studies should aim to include more diverse samples, these criteria are common in research with this population [75-77].

We demonstrated that the *Effectivate* SGCTA is sufficiently powerful to induce changes, even in cognitively healthy older adults, and that the E-WINDMIL test is sufficiently sensitive to detect such changes. Our preliminary results are the first step, suggesting the ability of the SGCTA *Effectivate* to engender far cognitive transfer. Future studies should also try to relate our results to other more traditional cognitive measures and questionnaires tapping users' subjective evaluation of their quality of life.

Summary and Implications

This exploratory study presents an early foray into the potential of speech processing in adverse conditions as a far-transfer gauge of SGCTAs. This is in line with previous studies that used gamification in cognitive decline research [78-81]. Results present a preliminary indicator of the SGCTA *Effectivate's* potential to engender such far transfer from visual cognitive training to auditory speech processing after only a few weeks of training. Following training, older adults were better able to differentiate between the spoken target word and its sound-sharing competitor under adverse conditions (noise and

XSL•FO
RenderX

digit memorization). We suggest that this change in performance represents a real-world improvement in a daily task that is directly related to successful aging. Thus, it shows the potential of the training to have a significant impact on the user's daily life. We advocate that cognitive training should showcase evidence-based improvement in daily far-transfer tasks that can change the user's quality of life, as opposed to merely showing changes in traditional pen-and-paper cognitive measures. As serious games are a means of improving performance in other tasks, games developed to the highest standards should seek out far-transfer validation methods. We hope that the increased demand for far-transfer metrics will bolster research efforts within the academic community to develop new far-transfer gauges of cognitive ability and call on serious game developers to adopt far-transfer metrics, such as E-WINDMIL, into their gauges for validity and success.

This study investigated aging through the lens of speech processing, a novel vantage point, which can illuminate interconnected attentional mechanisms known to be affected by aging. Most importantly, speech processing is an essential daily task performed across social interactions, leisure, and employment [72]. Impaired speech processing may have severe implications for older adults across all aspects of life. Therefore, we encourage adding tests of speech processing, especially in adverse conditions, to the arsenal of tools used to test the efficacy of EF training in aging. Furthermore, we suggest paying attention to speech processing in aging when considering accessibility and inclusion in serious game design.

In addition to being a novel and important test metric for aging, real-time speech processing metrics may also prove beneficial to testing other populations such as children with the neurodevelopmental disorder, attention-deficit/hyperactivity disorder (ADHD). As the most prevalent neurodevelopmental disorder in children, ADHD is associated with lifelong impairment, with symptoms reflecting a deficit in EFs such as inhibitory control, attentional regulation, and WM [82,83]. Given ADHD's high prevalence and detrimental effect on the quality of life and well-being, many serious games are being developed to train EFs in ADHD. As is the case with SGCTAs, there is much debate in the literature regarding their efficacy [84]. Expanding on our findings, we suggest further exploration using E-WINDMIL to test the far-transfer efficacy of serious computerized games designed for children and adults with ADHD along with other promising populations that could benefit. We hope that the creation and use of universally accepted far-transfer metrics will determine gold standard serious games that will help us prolong cognitive functions and improve well-being with age and throughout life.

Acknowledgments

This study was partially supported by a grant from *Effectivate* (Acerar Ltd). *Effectivate* was not involved in the study design; collection, analysis, and interpretation of data; writing of this paper; or decision to submit it for publication. The authors had full liberty in all of these aspects. The authors wish to acknowledge the assistance of the following students and staff in gathering the data: Maya Mentzel, Dana Yehezkel, and Tali Livne.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was not registered as the local Institutional Review Board did not believe it was needed. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low, the study was considered exploratory, and the authors' Institutional Review Board provided a letter of approval which recommended that registration not be mandated for this study. However, it should be noted such exceptions are uncommon and readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2969 KB - games_v10i3e32297_app1.pdf]

References

1. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2019-07-19]
2. Eggleston KN, Fuchs VR. The new demographic transition: most gains in life expectancy now realized late in life. J Econ Perspect 2012 Aug 01;26(3):137-156. [doi: 10.1257/jep.26.3.137] [Medline: 25076810]
3. The Handbook of Aging and Cognition. New York: Psychology Press; 2007.
4. Ben-David BM, Schneider BA. A sensory origin for color-word Stroop effects in aging: a meta-analysis. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2009 Sep 02;16(5):505-534. [doi: 10.1080/13825580902855862] [Medline: 19479479]

XSL•FO
RenderX

5. Ben-David BM, Schneider BA. A sensory origin for color-word stroop effects in aging: simulating age-related changes in color-vision mimics age-related changes in Stroop. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2010 Nov 08;17(6):730-746. [doi: 10.1080/13825585.2010.510553] [Medline: 21058053]
6. Ageism and neuropsychological tests. In: Contemporary Perspectives on Ageism. Cham: Springer; 2018.
7. Ben-David BM, Erel H, Goy H, Schneider BA. "Older is always better": age-related differences in vocabulary scores across 16 years. Psychol Aging 2015 Dec;30(4):856-862. [doi: 10.1037/pag0000051] [Medline: 26652725]
8. Morley JE. The top 10 hot topics in aging. J Gerontol A Biol Sci Med Sci 2004 Jan 01;59(1):24-33. [doi: 10.1093/gerona/59.1.m24] [Medline: 14718483]
9. Rowe JW, Kahn RL. Successful aging 2.0: conceptual expansions for the 21st Century. J Gerontol B Psychol Sci Soc Sci 2015 Jul 15;70(4):593-596. [doi: 10.1093/geronb/gbv025] [Medline: 25878054]
10. Ear and Hearing Care Planning and Monitoring of National Strategies. Geneva: World Health Organization; 2015.
11. Active ageing : a policy framework. World Health Organization. URL: https://apps.who.int/iris/handle/10665/67215 [accessed 2015-05-26]
12. Stanford Center on longevity letter. Cognitive Training Data. URL: https://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/stanford-letter/ [accessed 2022-07-10]
13. Cognitive training data response letter. Cognitive Training Data. URL: https://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter/ [accessed 2022-07-10]
14. Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, et al. Do "Brain-Training" programs work? Psychol Sci Public Interest 2016 Oct;17(3):103-186. [doi: 10.1177/1529100616661983] [Medline: 27697851]
15. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec;390(10113):2673-2734. [doi: 10.1016/s0140-6736(17)31363-6]
16. Effects of senescent changes in audition and cognition on spoken language comprehension. In: The Aging Auditory System. New York: Springer; 2010.
17. Davis A, McMahon CM, Pichora-Fuller KM, Russ S, Lin F, Olusanya BO, et al. Aging and hearing health: the life-course approach. Gerontologist 2016 Apr;56 Suppl 2:S256-S267. [doi: 10.1093/geront/gnw033] [Medline: 26994265]
18. Karpa MJ, Gopinath B, Beath K, Rochtchina E, Cumming RG, Wang JJ, et al. Associations between hearing impairment and mortality risk in older persons: the Blue Mountains Hearing Study. Ann Epidemiol 2010 Jun;20(6):452-459. [doi: 10.1016/j.annepidem.2010.03.011] [Medline: 20470972]
19. Pichora-Fuller M, Mick P, Reed M. Hearing, cognition, and healthy aging: social and public health implications of the links between age-related declines in hearing and cognition. Semin Hear 2015 Aug 9;36(3):122-139. [doi: 10.1055/s-0035-1555116] [Medline: 27516713]
20. Rönnberg J, Holmer E, Rudner M. Cognitive hearing science and ease of language understanding. Int J Audiol 2019 May 03;58(5):247-261. [doi: 10.1080/14992027.2018.1551631] [Medline: 30714435]
21. Cognition and hearing – you can't test one with the other!. Audiology Feature. URL: https://www.entandaudiologynews.com/media/23455/entma21-ben-david.pdf [accessed 2021-03-29]
22. van Tilburg TG, Steinmetz S, Stolte E, van der Roest H, de Vries DH. Loneliness and mental health during the COVID-19 pandemic: a study among Dutch older adults. J Gerontol B Psychol Sci Soc Sci 2021 Aug 13;76(7):e249-e255. [doi: 10.1093/geronb/gbaa111] [Medline: 32756931]
23. Palgi Y, Bergman YS, Ben-David B, Bodner E. No psychological vaccination: vaccine hesitancy is associated with negative psychiatric outcomes among Israelis who received COVID-19 vaccination. J Affect Disord 2021 May 15;287:352-353. [doi: 10.1016/j.jad.2021.03.064] [Medline: 33819733]
24. Keisari S, Palgi Y, Ring L, Folkman A, Ben-David BM. "Post-lockdown depression": adaptation difficulties, depressive symptoms, and the role of positive solitude when returning to routine after the lifting of nation-wide COVID-19 social restrictions. Front Psychiatry 2022 Mar 11;13:838903 [FREE Full text] [doi: 10.3389/fpsyt.2022.838903] [Medline: 35360132]
25. D'cruz M, Banerjee D. 'An invisible human rights crisis': the marginalization of older adults during the COVID-19 pandemic - An advocacy review. Psychiatry Res 2020 Oct;292:113369. [doi: 10.1016/j.psychres.2020.113369] [Medline: 32795754]
26. Palgi Y, Shrira A, Ring L, Bodner E, Avidor S, Bergman Y, et al. The loneliness pandemic: loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. J Affect Disord 2020 Oct 01;275:109-111. [doi: 10.1016/j.jad.2020.06.036] [Medline: 32658811]
27. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev 1996 Jul;103(3):403-428. [doi: 10.1037/0033-295x.103.3.403] [Medline: 8759042]
28. Salthouse TA. The aging of working memory. Neuropsychology 1994;8(4):535-543. [doi: 10.1037/0894-4105.8.4.535]
29. Working memory. In: Psychology of Learning and Motivation. Amsterdam, Netherlands: Elsevier Science; 1974.
30. McCabe DP, Roediger HL, McDaniel MA, Balota DA, Hambrick DZ. The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. Neuropsychology 2010 Mar;24(2):222-243. [doi: 10.1037/a0017619] [Medline: 20230116]
31. Postle B. Working memory as an emergent property of the mind and brain. Neuroscience 2006 Apr 28;139(1):23-38. [doi: 10.1016/j.neuroscience.2005.06.005] [Medline: 16324795]

XSL•FO
RenderX

32. Gordon-Salant S, Yeni-Komshian GH, Fitzgibbons PJ, Barrett J. Age-related differences in identification and discrimination of temporal cues in speech segments. J Acoust Soc Am 2006 Apr;119(4):2455-2466. [doi: 10.1121/1.2171527] [Medline: 16642858]
33. The contribution of auditory and cognitive factors to intelligibility of words and sentences in noise. In: Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing. Cham: Springer; 2016.
34. Benichov J, Cox LC, Tun PA, Wingfield A. Word recognition within a linguistic context: effects of age, hearing acuity, verbal ability, and cognitive function. Ear Hear 2012;33(2):250-256. [doi: 10.1097/AUD.0b013e31822f680f] [Medline: 21918453]
35. Ezzatian P, Li L, Pichora-Fuller K, Schneider BA. Delayed stream segregation in older adults: more than just informational masking. Ear Hear 2015;36(4):482-484. [doi: 10.1097/AUD.0000000000000139] [Medline: 25587669]
36. Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. Nature 2005 Nov 24;438(7067):500-503. [doi: 10.1038/nature04171] [Medline: 16306992]
37. Janse E. A non-auditory measure of interference predicts distraction by competing speech in older adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2012 Nov;19(6):741-758. [doi: 10.1080/13825585.2011.652590] [Medline: 22293017]
38. Salthouse TA. Resource-reduction interpretations of cognitive aging. Development Rev 1988 Sep;8(3):238-272. [doi: 10.1016/0273-2297(88)90006-8]
39. Lash A, Wingfield A. A Bruner-Potter effect in audition? Spoken word recognition in adult aging. Psychol Aging 2014 Dec;29(4):907-912. [doi: 10.1037/a0037829] [Medline: 25244463]
40. Lash A, Rogers CS, Zoller A, Wingfield A. Expectation and entropy in spoken word recognition: effects of age and hearing acuity. Exper Aging Res 2013 May;39(3):235-253. [doi: 10.1080/0361073x.2013.779175]
41. Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BW, Humes LE, et al. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). Ear Hear 2016;37 Suppl 1:5S-27S. [doi: 10.1097/AUD.0000000000000312] [Medline: 27355771]
42. Cohen G, Faulkner D. Word recognition: age differences in contextual facilitation effects. Br J Psychol 1983 May;74(Pt 2):239-251. [doi: 10.1111/j.2044-8295.1983.tb01860.x] [Medline: 6883014]
43. Dubno JR, Ahlstrom JB, Horwitz AR. Use of context by young and aged adults with normal hearing. J Acoust Soc Am 2000 Jan;107(1):538-546. [doi: 10.1121/1.428322] [Medline: 10641662]
44. Rönnberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, et al. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front Syst Neurosci 2013;7:31 [FREE Full text] [doi: 10.3389/fnsys.2013.00031] [Medline: 23874273]
45. Working memory, comprehension, and aging: a review and a new view. In: The Psychology of Learning and Motivation. Amsterdam, Netherlands: Elsevier Science; 2003.
46. Harel-Arbeli T, Wingfield A, Palgi Y, Ben-David BM. Age-related differences in the online processing of spoken semantic context and the effect of semantic competition: evidence from eye gaze. J Speech Lang Hear Res 2021 Feb 17;64(2):315-327. [doi: 10.1044/2020_jslhr-20-00142]
47. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug;396(10248):413-446. [doi: 10.1016/s0140-6736(20)30367-6]
48. Lin FR, Albert M. Hearing loss and dementia - who is listening? Aging Ment Health 2014;18(6):671-673 [FREE Full text] [doi: 10.1080/13607863.2014.915924] [Medline: 24875093]
49. Deal JA, Goman AM, Albert MS, Arnold ML, Burgard S, Chisolm T, et al. Hearing treatment for reducing cognitive decline: design and methods of the Aging and Cognitive Health Evaluation in Elders randomized controlled trial. Alzheimers Dement (N Y) 2018 Oct 05;4(1):499-507 [FREE Full text] [doi: 10.1016/j.trci.2018.08.007] [Medline: 30364572]
50. Pratt J, Dodd M, Welsh T. Growing older does not always mean moving slower: examining aging and the saccadic motor system. J Motor Behav 2006 Sep;38(5):373-382. [doi: 10.3200/jmbr.38.5.373-382]
51. Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science 1995 Jun 16;268(5217):1632-1634. [doi: 10.1126/science.7777863] [Medline: 7777863]
52. Nitsan G, Wingfield A, Lavie L, Ben-David BM. Differences in working memory capacity affect online spoken word recognition: evidence from eye movements. Trends Hear 2019 Apr 23;23:2331216519839624 [FREE Full text] [doi: 10.1177/2331216519839624] [Medline: 31010398]
53. Nitsan G, Banai K, Ben-David BM. One size does not fit all: examining the effects of working memory capacity on spoken word recognition in older adults using eye tracking. Front Psychol 2022 Apr 11;13:841466 [FREE Full text] [doi: 10.3389/fpsyg.2022.841466] [Medline: 35478743]
54. Baharav S, Nitsan G, Ben-David BM. Commentary: Working Memory Load Affects Processing Time in Spoken Word Recognition: Test Retest Reliability of the E-WINDMIL Eyetracking Paradigm. Front Neurosci 2021;15:663930 [FREE Full text] [doi: 10.3389/fnins.2021.663930] [Medline: 34177448]
55. Ben-David BM, Tse VY, Schneider BA. Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? Hear Res 2012 Aug;290(1-2):55-63. [doi: 10.1016/j.heares.2012.04.022] [Medline: 22609772]

XSL•FO
RenderX

56. Hadar B, Skrzypek JE, Wingfield A, Ben-David BM. Working memory load affects processing time in spoken word recognition: evidence from eye-movements. Front Neurosci 2016 May 19;10:221. [doi: 10.3389/fnins.2016.00221] [Medline: 27242424]
57. Salverda AP, Tanenhaus MK. The visual world paradigm. In: Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide. Hoboken, New Jersey, United States: Wiley; 2017.
58. Kee J, Knuth M, Lahey JN, Palma MA. Does eye-tracking have an effect on economic behavior? PLoS One 2021 Aug 5;16(8):e0254867 [FREE Full text] [doi: 10.1371/journal.pone.0254867] [Medline: 34351921]
59. Rashbass C. The relationship between saccadic and smooth tracking eye movements. J Physiol 1961 Dec;159:326-338 [FREE Full text] [doi: 10.1113/jphysiol.1961.sp006811] [Medline: 14490422]
60. Ben-David BM, Chambers CG, Daneman M, Pichora-Fuller MK, Reingold EM, Schneider BA. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. J Speech Lang Hear Res 2011 Feb;54(1):243-262. [doi: 10.1044/1092-4388(2010/09-0233)]
61. Arnold JE, Fagnano M, Tanenhaus MK. Disfluencies signal Theee, Um, New Information. J Psycholinguist Res 2003;32:25-36. [doi: 10.1023/A:1021980931292]
62. Brown-Schmidt S. The role of executive function in perspective taking during online language comprehension. Psychonomic Bulletin Rev 2009 Oct;16(5):893-900. [doi: 10.3758/pbr.16.5.893]
63. Kaiser E, Trueswell JC. Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. Language Cognit Processes 2008 Aug;23(5):709-748. [doi: 10.1080/01690960701771220]
64. Lin FR. Hearing loss and cognition among older adults in the United States. J Gerontol A Biol Sci Med Sci 2011 Oct 18;66(10):1131-1136 [FREE Full text] [doi: 10.1093/gerona/glr115] [Medline: 21768501]
65. Ayasse ND, Lash A, Wingfield A. Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. Front Aging Neurosci 2016 Jan 10;8:329 [FREE Full text] [doi: 10.3389/fnagi.2016.00329] [Medline: 28119598]
66. Huettig F, Janse E. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. Language Cognition Neuroscience 2015 Sep 21;31(1):80-93. [doi: 10.1080/23273798.2015.1047459]
67. Rönnberg J, Rudner M, Foo C, Lunner T. Cognition counts: a working memory system for ease of language understanding (ELU). Int J Audiol 2008 Nov 07;47 Suppl 2(sup2):S99-105. [doi: 10.1080/14992020802301167] [Medline: 19012117]
68. Helfer KS, Staub A. Competing speech perception in older and younger adults: behavioral and eye-movement evidence. Ear Hear 2014;35(2):161-170 [FREE Full text] [doi: 10.1097/AUD.0b013e3182a830cf] [Medline: 24351611]
69. Ben-David BM, Avivi-Reich M, Schneider BA. Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? Hear Res 2016 Nov;341:9-18. [doi: 10.1016/j.heares.2016.07.016] [Medline: 27496539]
70. Snyder JS, Alain C. Toward a neurophysiological theory of auditory stream segregation. Psychol Bull 2007 Sep;133(5):780-799. [doi: 10.1037/0033-2909.133.5.780] [Medline: 17723030]
71. Snyder JS, Alain C, Picton TW. Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 2006 Jan;18(1):1-13. [doi: 10.1162/089892906775250021] [Medline: 16417678]
72. Heinrich A, Gagne JP, Viljanen A, Levy DA, Schneider BA. Effective communication as a fundamental aspect of active aging and well-being: paying attention to the challenges older adults face in noisy environments. Social Inquiry Well-Being 2016;2(1). [doi: 10.13165/SIIW-16-2-1-05]
73. The physiology of cochlear presbycusis. In: The Aging Auditory System. Springer Handbook of Auditory Research. New York: Springer; 2010.
74. Revill KP, Spieler DH. The effect of lexical frequency on spoken word recognition in young and older listeners. Psychol Aging 2012 Mar;27(1):80-87 [FREE Full text] [doi: 10.1037/a0024113] [Medline: 21707175]
75. Nagar S, Mikulincer M, Nitsan G, Ben-David BM. Safe and sound: the effects of experimentally priming the sense of attachment security on pure-tone audiometric thresholds among young and older adults. Psychol Sci 2022 Mar 17;33(3):424-432. [doi: 10.1177/09567976211042008] [Medline: 35175871]
76. Keisari S, Feniger-Schaal R, Palgi Y, Golland Y, Gesser-Edelsburg A, Ben-David B. Synchrony in old age: playing the mirror game improves cognitive performance. Clin Gerontol 2022 Aug 06;45(2):312-326. [doi: 10.1080/07317115.2020.1799131] [Medline: 32762289]
77. Ben-David BM, Icht M. The effect of practice and visual feedback on oral-diadochokinetic rates for younger and older adults. Lang Speech 2018 Mar 14;61(1):113-134. [doi: 10.1177/0023830917708808] [Medline: 28610466]
78. Frost E, Porat T, Malhotra P, Picinali L. A novel auditory-cognitive training app for delaying or preventing the onset of dementia: participatory design with stakeholders. JMIR Hum Factors 2020 Sep 30;7(3):e19880 [FREE Full text] [doi: 10.2196/19880] [Medline: 32996884]
79. Masurovsky A. Controlling for placebo effects in computerized cognitive training studies with healthy older adults from 2016-2018: systematic review. JMIR Serious Games 2020 Jun 26;8(2):e14030 [FREE Full text] [doi: 10.2196/14030] [Medline: 32589159]

XSL•FO
RenderX

80. Israsena P, Jirayucharoensak S, Hemrungrojn S, Pan-Ngum S. Brain exercising games with consumer-grade single-channel electroencephalogram neurofeedback: pre-post intervention study. JMIR Serious Games 2021 Jun 15;9(2):e26872 [FREE Full text] [doi: 10.2196/26872] [Medline: 34128816]
81. Krebs C, Falkner M, Niklaus J, Persello L, Klöppel S, Nef T, et al. Application of eye tracking in puzzle games for adjunct cognitive markers: pilot observational study in older adults. JMIR Serious Games 2021 Mar 22;9(1):e24151 [FREE Full text] [doi: 10.2196/24151] [Medline: 33749607]
82. Castellanos FX, Sonuga-Barke EJ, Milham MP, Tannock R. Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cogn Sci 2006 Mar;10(3):117-123. [doi: 10.1016/j.tics.2006.01.011] [Medline: 16460990]
83. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [doi: 10.1037/0033-2909.121.1.65] [Medline: 9000892]
84. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015 Apr 6;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
E-WINDMIL: Eye tracking of Word Identification in Noise Under Memory Increased Load
EF: executive function
SGCTA: serious game for cognitive training in aging
WM: working memory

Edited by N Zary; submitted 24.07.21; peer-reviewed by A Richert-Kaźmierska, M Aksoy; comments to author 03.12.21; revised version received 21.04.22; accepted 28.04.22; published 28.07.22.

Please cite as:
Nitsan G, Baharav S, Tal-Shir D, Shakuf V, Ben-David BM
Speech Processing as a Far-Transfer Gauge of Serious Games for Cognitive Training in Aging: Randomized Controlled Trial of Web-Based Effectivate Training
JMIR Serious Games 2022;10(3):e32297
URL: https://games.jmir.org/2022/3/e32297
doi:10.2196/32297
PMID:35900825

©Gal Nitsan, Shai Baharav, Dalith Tal-Shir, Vered Shakuf, Boaz M Ben-David. Originally published in JMIR Serious Games (https://games.jmir.org), 28.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual ER, a Serious Game for Interprofessional Education to Enhance Teamwork in Medical and Nursing Undergraduates: Development and Evaluation Study

Janet Yuen-Ha Wong[1], PhD, RN; Joanna Ko[1], DN; Sujin Nam[1], MN; Tyrone Kwok[2], PhD; Sheila Lam[1], MN; Jessica Cheuk[1], MSc; Maggie Chan[1], DHSc; Veronica Lam[1], DN; Gordon T C Wong[3], MD; Zoe L H Ng[1], MEd; Abraham Ka-Chung Wai[4,5], MBChB, MSc, JD

[1]School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong

[2]Technology-Enriched Learning Initiative, The University of Hong Kong, Hong Kong, Hong Kong

[3]Department of Anaesthesiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong

[4]Department of Emergency Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong

[5]Accident and Emergency Department, The University of Hong Kong–Shenzhen Hospital, Shenzhen, China

Corresponding Author:
Abraham Ka-Chung Wai, MBChB, MSc, JD
Department of Emergency Medicine
School of Clinical Medicine, Li Ka Shing Faculty of Medicine
The University of Hong Kong
5 Sassoon Road, Pokfulam
Hong Kong
Hong Kong
Phone: 852 39179175
Email: awai@hku.hk

Abstract

Background: Engaging students in interprofessional education for higher order thinking and collaborative problem-solving skills is challenging. This study reports the development of Virtual ER, a serious game played on a virtual platform, and how it can be an innovative way for delivering interprofessional education to medical and nursing undergraduates.

Objective: We report the development of a serious online game, Virtual ER, and evaluate its effect on teamwork enhancement and clinical competence. We also explore if Virtual ER can be an effective pedagogical tool to engage medical and nursing students with different learning styles.

Methods: Virtual ER is a custom-made, learning outcome–driven, case-based web app. We developed a game performance scoring system with specific mechanisms to enhance serious gaming elements. Sixty-two students were recruited from our medical and nursing programs. They played the games in teams of 4 or 5, followed by an instructor-led debriefing for concept consolidation. Teamwork attitudes, as measured by the Human Factors Attitude Survey, were compared before and after the game. Learning style was measured with a modified Honey and Mumford learning style questionnaire.

Results: Students were satisfied with Virtual ER (mean satisfaction score 5.44, SD 0.95, of a possible 7). Overall, Virtual ER enhanced teamwork attitude by 3.02 points (95% CI 1.15-4.88, *P*=.002). Students with higher scores as activists (estimate 9.09, 95% CI 5.17-13.02, *P*<.001) and pragmatists (estimate 5.69, 95% CI 1.18-10.20, *P*=.01) had a significantly higher degree of teamwork attitude enhancement, while students with higher scores as theorists and reflectors did not demonstrate significant changes. However, there was no difference in game performance scores between students with different learning styles.

Conclusions: There was considerable teamwork enhancement after playing Virtual ER for interprofessional education, in particular for students who had activist or pragmatist learning styles. Serious online games have potential in interprofessional education for the development of 21st century life skills. Our findings also suggest that Virtual ER for interprofessional education delivery could be expanded locally and globally.

XSL•FO
RenderX

(JMIR Serious Games 2022;10(3):e35269) doi:10.2196/35269

KEYWORDS

game; interprofessional education; teamwork; learning style; emergency medicine; emergency nursing

Introduction

The emergency room (ER) is a unique clinical environment in which effective teamwork between doctors and nurses is necessary for prioritization of patient care, complex clinical decision-making, and safe yet efficient clinical practice [1]. However, these professionals are normally trained in their own professional silos with separate curricula and assessment schemes. The World Health Organization recommends interprofessional education (IPE) [2], which is increasingly recognized as a paramount tool to purposefully bring students from different disciplines together to learn teamwork, communications skills, and patient safety culture for collaboration in their future clinical practice.

IPE is challenging because of its demands for physical space, faculty time, timetable alignment, and student engagement [3]. However, its value is reflected in the successful enhancement of teamwork, as evidenced by many successful cases in previous reports [4-6]. With the teaching disruptions caused by the COVID-19 pandemic, the development of alternative methods of IPE delivery through online platforms is necessary to meet unmet demands.

Engaging students with collaborative problem-solving skills through higher-order thinking in online IPE is a challenge [7]. Online serious gaming is an innovative delivery mechanism for IPE. Unlike gamification, serious gaming should align with learning goals and drive students to engage in the game due to intrinsic factors, such as autonomy, relatedness, and competence, instead of extrinsic factors, such as points and badges [8]. There are diverse definitions of serious games in the literature. Bedwell and colleagues [9] empirically identified, in a systematic review, many attributes of serious games, including action language, assessment, conflicts and challenge, control, environment, game fiction, human interaction, immersion, and rules and goals. Indeed, the growing popularity of serious gaming in clinical education has been noted since 2010 [10,11]. Regarding teamwork and communications enhancement, there is also evidence showing that game-based surgical education training, with face-detection and head-tracking technology, can enhance nonverbal communications and facilitate teamwork [12]. There is also evidence that single-player screen-based teamwork training is feasible, and that it does not affect in-game performance [13]. However, there is a dearth of evaluation studies of serious online games for IPE to enhance teamwork among undergraduate students. Teamwork with members from different disciplines is particularly important in the ER to enhance diagnostic accuracy and efficiency and prevent medical errors in transitions of care. Therefore, IPE in undergraduate education is known to be essential for enabling an effective, collaborative practice eHealth-ready workforce for health care settings [14].

Clinical education today emphasizes student-centered approaches in teaching and learning. Although evidence supports the benefits of IPE, students' attitudes toward IPE vary because of their highly diverse backgrounds, compounded by their different professional programs, different levels of clinical exposure, and different learning styles [15]. It is understood that different learning styles may affect student learning outcomes. However, there is a gap in evidence to understand how serious games in clinical IPE can be effective for students with different learning styles. Kolb's learning cycle identified 4 learning styles: reflectors, theorists, activists, and pragmatists [16]. It has also been noted that face-to-face simulation-based activities are most effective for students who have a balanced learning style, including those who are theorists; in other words, they are effective for students who learn through abstract thinking, reflection, and carefully looking into problems from multiple perspectives [17]. On the other hand, game-based activities are most effective in activists and reflectors, who prefer concrete experience and reflective observation [18]. Therefore, awareness of learning styles will be helpful for clinical educators to modify and personalize gamification elements for effective strategies for student engagement in IPE.

In this study, we developed Virtual ER, an online serious game set in the context of an emergency room with specific learning goals. We evaluated the effect of Virtual ER on teamwork enhancement via a pilot study. Although we anticipated that this serious game would be an effective pedagogical tool to engage students in medicine and nursing, we were also interested in whether learning style, based on Kolb's learning cycle [16], would be a key driving factor underlying the impact of serious games in IPE. We hypothesized that students who are activists and pragmatists would have greater score increases for teamwork attitude, as measured by the Human Factors Attitude Survey, compared to students who are reflectors or theorists (hypothesis 1). Moreover, we hypothesized that students who are activists and pragmatists would have higher scores for clinical competence compared with those who are reflectors and theorists (hypothesis 2).

We also examined whether students' prior basic life support or advanced cardiac life training affected the primary outcomes and game performance scores. Our findings may lead to recommendations for pedagogical designs and best practices that are relevant to serious games for IPE with student-centered approaches.

Methods

Serious Game Development

Design Philosophy

Using constructivism as the underlying learning philosophy, our Virtual ER game provides a platform for cognitive, socio-cultural, and collaborative interactions for student learning

XSL•FO
RenderX

[19]. Students as players are required to analyze scenarios and situations in the game and collaborate with others to select correct answers and demonstrate clinical competency.

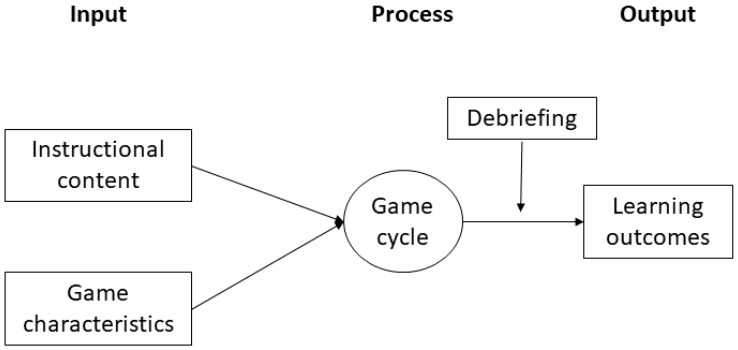
Design Framework

Virtual ER is a serious game designed based on the input-process-output (IPO) model, which is a tacit model of learning used in many studies [20-22]. Figure 1 illustrates the IPO model. First, the instructional program was designed to meet the learning goals and incorporated certain game features to facilitate students' learning according to the goals. The learning goals included competency in history taking, physical assessment skills, diagnostic skills, alertness, treatment skills, clinical procedural skills, prioritization skills during emergency situations, professionalism, responsibility, and a caring attitude. Table 1 describes specific learning goals in a specific scenario used in the game. The scenario was introduced to the players as follows:

> *A 66-year-old man, Peter, was admitted to Accident & Emergency Department this morning due to intermittent chest pain for the last 24 hours. He was a chronic smoker and was previously diagnosed with hypertension and hyperlipidemia. Without any prior symptoms, he started to develop shortness of breath and decreased exercise tolerance in these few months. One hour ago, he suffered from severe chest pain with radiation to left shoulder. One tablet of glyceryl trinitrate was taken but effect was minimal. He then sought emergency medical care this morning due to pain persistence. Aspirin was given in the ambulance during transport after ruling out NKDA [no known drug allergies]. Peter's wife, Jenny, accompanied upon admission. In this game, you are the nurse or physician responsible for Peter, who requires emergency care.*

Second, the game enhances student communication and collaborative efforts by providing feedback on their engagement in the game and increasing their motivation to play. Engagement in game play, feedback from the game, and, most importantly, a debriefing session led to the achievement of learning goals and specific learning outcomes.

Figure 1. Illustration of the input-process-output model.



XSL•FO
RenderX

Table 1. Specific learning goals in a specific scenario used in the game.

Game stage	Learning goals	Game description	Student tasks (examples)	
			Medical students	Nursing students
Triage station	Competency in history taking skills: gathering relevant and accurate health information via history taking	Categorizing patients via the triage system	• Be alert to the criteria of triage categorization • Perform comprehensive assessment and history taking upon arrival • Arrange appropriate investigation to rule out possible diagnoses	• Be alert to patient severity and perform immediate medical management • Assess pain level and determine if immediate treatment should be given at triage • Obtain baseline vital signs upon patient arrival
Emergency cubicle	Competency in physical assessment skills, diagnostic skills, alertness, and treatment skills	Assessing patients' condition in the emergency cubicle	• Be alert to any signs of deterioration and perform prompt medical management • Cooperate and communicate with team members to provide emergency care to the patient	• Monitor patients closely and perform reassessments to obtain data for guiding further nursing management • Be alert for any signs of deterioration and inform doctors if necessary
Resuscitation room	Competency in clinical procedural skills and prioritization skills during emergency situations	Providing and prioritizing appropriate treatment in an emergency situation	• Perform emergency care efficiently according to advanced cardiac life support protocols • Become familiarized with the use of different emergency medications during cardiopulmonary resuscitation • Appropriately intubate patients in respiratory distress	• Perform emergency care efficiently according to advanced cardiac life support protocols • Cooperate and communicate with doctors to ensure emergency management can be promptly delivered. • Assist doctors to perform intubation for patients with respiratory distress
Discharge to cardiac care unit	Competency in professionalism, responsibility, and a caring attitude	Preparing for a case transferal in emergency room	• Consult corresponding department for taking over stabilized patients	• Communicate with corresponding department about the condition of patients to guarantee continuous care can be arranged • Prepare case transfer with valid documentation and proper functioning of medical support devices

Instructional Content

Virtual ER begins with a virtual patient attending the emergency room on a Monday morning at 10 AM. There are 4 or 5 students playing doctor or nurse avatars with roles in triage, initial assessment, monitoring, resuscitation, and transfer of care. Virtual ER allows students to assess patients through their medical history, a physical examination (including vital signs), investigation (eg, electrocardiography, X-rays, CT scans, blood tests, and point-of-care tests), treatment (eg, prescriptions, drug administration, establishment of intravenous access, oxygen therapy, suturing, and wound care), and nursing processes, including caring (eg, phone calls, talking to patients to provide comfort, talking to relatives), charting, discharge, and transfer to different settings or rooms (Figure 2). The virtual patient's condition deteriorates during the course of care and progresses to a need for resuscitation. Student players are required to provide urgent care to the patient according to different priorities. When we set up the challenges for students, some questions were designed particularly for medical students and some questions for nursing students. Performance scores were allocated to the teams to complete decision-making or behavioral tasks accurately, safely, and properly. Table 1 shows the students' responsibilities in managing the virtual patient.

XSL•FO
RenderX

Figure 2. Screenshot of Virtual ER. The scenario starts at the triage center. Students can go into different rooms according to the scenario, including the resuscitation room.



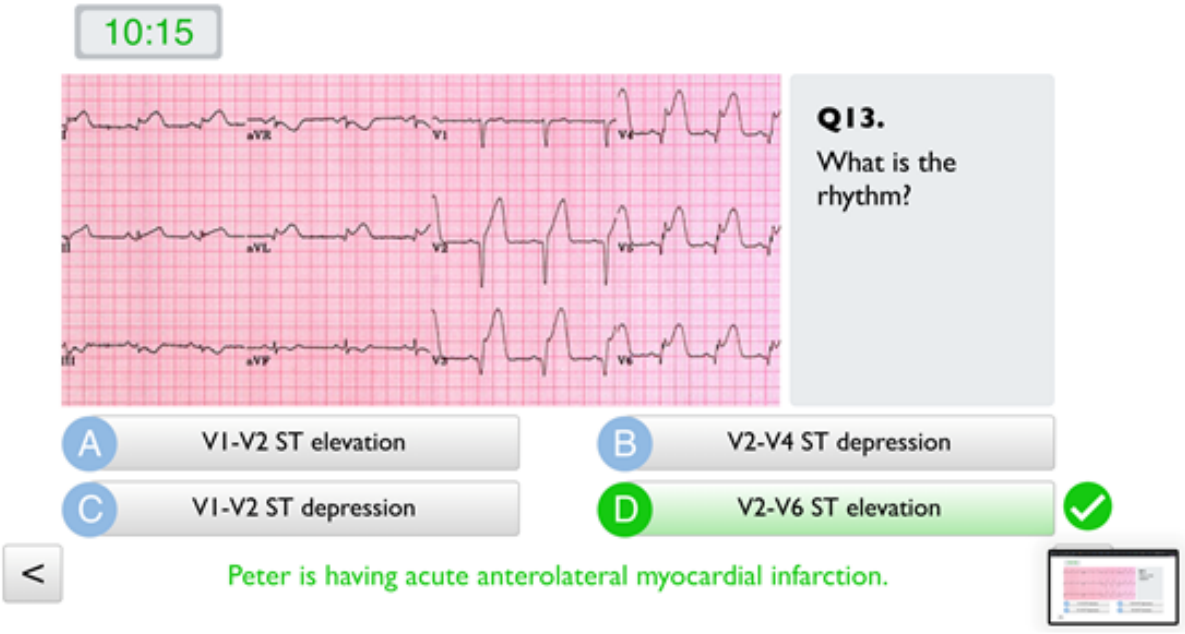
Game Characteristics and Performance Scoring System

Virtual ER is a custom-made web app built with Tumult Hype (Tumult Inc), a tool for creating interactive web content. The performance scoring system used three mechanisms to enhance student learning: (1) the students' first response to each question (ie, their selected actions and choices in the game) was stored in the system and allocated to the team. This facilitated teachers' performance analysis and subsequent debriefing by the end of the class. It also left the students unaware of their performance during game play, leaving them to discover their number of mistakes at the end of the game [23]. Figure 3 shows a question about electrocardiogram results. (2) The students were allowed to go back to prior questions and select other actions or choices, but this did not lead to any change in the score assigned based on the already attempted question. This allowed students to think carefully before answering questions and enhanced face-to-face interpersonal discussion among team members [24]. This rule was emphasized at the beginning of the game. (3) Incorrect actions and choices led to score deductions, to drive students to rationalize each decision and discourage answering randomly. This penalty system was designed to discourage undesirable behaviors in a learning environment [25].

The scoring system was implemented using JavaScript inside the Tumult Hype application. The students' performance data was stored on the Firebase cloud server (Google LLC). These data are downloadable for analysis.

XSL•FO
RenderX

Figure 3. Screenshot of Virtual ER. The question is about interpreting electrocardiogram results. ER: Emergency room. Q3: Question 3. ST: represents wave on an electrocardiogram. V1-V6: represents electrocardiogram leads.



Procedures

The learning goal of Virtual ER was to develop a sense of teamwork among medical and nursing students. Students were assigned into interdisciplinary groups with 4 to 5 students each. They accessed Virtual ER online simultaneously on laptops. Before the game started, all students were briefed on the educational aims and objectives and provided game orientation. Written informed consent for evaluation was then sought. The students were invited to complete a 10-minute pretest questionnaire. Instructions for downloading the online game were also provided. The students were given 30 minutes to play the game with their teammates. During the game, teachers encouraged the students to share their domain knowledge and to discuss the best clinical decisions for the virtual patient. All their decisions or behaviors during the game were evaluated systematically for the clinical competence they showed regarding patient management. At the end of the game, all the groups were invited to join the debriefing as a whole class. Figure 4 shows student peers discussing the Virtual ER game. Emergency medicine and nursing teachers cotaught during the debriefing, with the dynamics and experiences of different groups linked to teamwork and clinical competency in handling the virtual patient in the scenario. Upon completion of the debriefing, the students were invited to fill out a posttest questionnaire, which took around 10 minutes to complete. The whole session lasted for about 2 hours.

Figure 4. Images of student discussion while playing Virtual ER. ER: Emergency room.



XSL•FO
RenderX

Participants and Setting

All students in the emergency clerkship course and all nursing undergraduates in the emergency nursing-care course of a university were invited to participate via email and e-flyer. The pilot study took place between August 2021 and November 2021. Sixty-two final-year students participated. They were well equipped with necessary clinical knowledge and skills and had received relevant emergency care training prior to this study.

Evaluation Measurements

Attitudes Toward Teamwork (Primary Outcome)

We assessed attitudinal shifts to team behavior with the Human Factors Attitude Survey (HFAS), which was developed by the University of Texas and the National Aeronautics and Space Administration [26]. The HFAS has 23 items with good internal reliability (Cronbach α=.89) and has been tested in Hong Kong [27]. All questions were reviewed by expert doctors and nurses for content and face validity to ensure that the survey was subjectively viewed as covering the concepts it measured. Participants were asked to indicate their agreement with each question on a 5-point Likert scale (from 1, indicating "strongly disagree," to 5, "strongly agree"). Some examples of questions include the following: "my performance is not adversely affected by working with an inexperienced or less capable team member," "prior to the procedure, it is important for all team members to be familiar with the tasks and responsibilities of the other members of the team," and "my ability to detect adverse situations has a direct relationship to the quality of decisions I make." This scale was used in pre- and posttest questionnaires. The total score ranged between 23 and 115.

Clinical competencies were assessed for handling patients in the case scenarios. Students were provided with a performance score upon completion of a case scenario. Marks were given for correct answers in the game and marks were deducted for incorrect answers. The total possible score for each case was 41.

Learning Style

A modified version of the Honey and Mumford learning style questionnaire was used [28]. It contained 13 items with responses on a 5-point Likert scale (1 indicating "strongly disagree," 2 indicating "disagree," 3 indicating "neutral," 4 indicating "agree," and 5 indicating "strongly agree"). It was tested and found to have satisfactory Cronbach α values, ranging from α=.593 to α=.786 for the 4 different types of learners: reflectors, theorists, activists, and pragmatists. The higher the scores for a specific learning style, the more likely it was that the students adopted that learning style. Learning styles were assessed in the pretest questionnaire.

Demographics

Demographics included sex, study program, and clinical part-time job experience. Information on additional basic life support and advanced cardiovascular life support training was also gathered with the pretest questionnaire.

Virtual ER Game Satisfaction

One question was included in the posttest questionnaire to assess game satisfaction on a 7-point Likert scale, ranging from "not at all satisfied" to "extremely satisfied." The question was "How satisfied are you with the game and tutorial?"

Data Management and Analysis

Data were collected with an online questionnaire that used the Qualtrics platform (Qualtrics). Descriptive statistics were used to describe the students' characteristics. Quantitative data were entered into SPSS (IBM Co). Changes in teamwork attitude over time were assessed with a paired *t* test (2-tailed). We used linear mixed effect modeling (LMM) to evaluate whether students' learning styles and prior training in basic life support or advanced cardiovascular life support had any effect on changes in teamwork attitude. LMM has been used for the analysis of between-participant data, including both fixed effects and random effects [29]. This was helpful for the robustness of the data analysis, in particular, for evaluating the effects of differences in the study population in terms of learning styles and a priori training and clinical experience.

Ethical Approval

Ethical approval was obtained from the University of Hong Kong/Hospital Authority Hong Kong West Cluster Joint Institutional Review Board (UW 21-302). A signed online consent form was obtained before data collection. During the informed consent process, an information sheet was provided that included information on the credentials and affiliations of the researchers and the purpose of the evaluation.

Results

Of 62 students, 34 (54%) were male and 47 (76%) were studying medicine. There were 38 students (61%) who had attended basic life support training and 35 (56%) students who had attended advanced cardiovascular life support training before joining the Virtual ER game workshop. There were 20 (32%) activists, 19 (31%) pragmatists, 4 (7%) reflectors, and 3 (5%) theorists. There were 16 students (26%) with a combination of 2, 3, or 4 styles. Mean scores on the Honey and Mumford questionnaire for the activists, theorists, reflectors, and pragmatists were 3.76 (SD 0.49), 3.47 (SD 0.52), 3.24 (SD 0.62), and 3.71 (SD 0.47), respectively.

Overall, Virtual ER enhanced teamwork attitudes. The pretest HFAS score was 91.26 (SD 8.75) and the posttest score was 94.27 (SD 10.46), representing a 3.02-point increase (95% CI 1.15-4.88, *P*=.002). We used a mixed effect model to further investigate the effect of learning style on teamwork attitude enhancement (Table 2 shows the results). As we hypothesized (hypothesis 1), students who had higher scores as activists (estimate 9.09, 95% CI 5.17-13.02, *P*<.001) and pragmatists (estimate 5.69, 95% CI 1.18-10.20, *P*=.01) had significantly higher teamwork attitude enhancement, while students with higher scores as theorists and reflectors showed no significant change. Regarding the Virtual ER game performance scores, there was no difference among students with different learning styles. Therefore, hypothesis 2 was not supported.

XSL•FO
RenderX

Students with prior experience in basic life support showed no significant difference in teamwork enhancement or Virtual ER game performance. However, students with prior advanced cardiovascular life support training showed a decrease in teamwork attitude between baseline and after playing the game (estimate –4.86, 95% CI –9.13 to –0.59, *P*=.03).

In general, students were satisfied with Virtual ER, with a game satisfaction score of mean 5.44 (SD 0.95) points of a possible 7. Of the 62 students, 2 (3%) rated their experience with the app "not satisfied," 6 (10%) rated it "neutral," 25 (40%) rated it "satisfied," 21 (34%) rated it "very satisfied," and 8 (13%) rated it "extremely satisfied."

Table 2. Effect of learning style on teamwork enhancement and performance scores.

	Teamwork enhancement			Virtual ER[a] game performance		
Outcomes	Estimate	95% CI	*P* value	Estimate	95% CI	*P* value
Learning style						
Activist	9.09	5.17 to 13.02	<.001	–5.75	14.77 to 3.26	.21
Theorist	1.81	–2.46 to 6.08	.41	–8.17	–16.52 to 0.17	.06
Reflector	2.49	–1.04 to 6.02	.16	–6.42	–13.42 to 0.58	.07
Pragmatist	5.69	1.18 to 10.20	.01	–5.03	–14.44 to 4.38	.29
Prior experience						
Basic life support	–2.84	–9.48 to 1.13	.12	–1.70	–10.77 to 7.36	.71
Advanced cardiovascular life support	–4.86	–9.13 to –0.59	.03	3.87	–4.99 to 12.73	.39

[a]ER: Emergency room.

Discussion

Principal Findings

Using a serious game approach for clinical education is an emerging field, and our study is among the first to report on the development of a custom-made, ad-hoc, collaborative game, Virtual ER, and evaluate it for its ability to enhance teamwork among medical and nursing undergraduate students. More importantly, we demonstrate that Virtual ER enhanced teamwork in students with a small but statistically significant effect, and that it was particularly and significantly effective in students who had more activist or pragmatist personalities. This finding supports our hypothesis 1. Compared with our previous study, which examined face-to-face interprofessional clinical simulation [5], Virtual ER was as effective at enhancing teamwork among medical and nursing students as clinical simulation. In addition, Virtual ER was widely accepted by the medical and nursing students and was rated highly for satisfaction. Apart from establishing the validity of serious games in interprofessional education, it was worthwhile to provide an enjoyable and motivating environment for students.

Using serious games for interprofessional learning outcomes is not new. Previous studies have used card or board games [30-33] to demonstrate positive outcomes for teamwork, communication, and interprofessional roles. Their results are consistent with our current findings on teamwork enhancement as a primary outcome. However, our study is innovative in terms of the pedagogical material that it included as part of the routine curriculum, giving students the opportunity to learn not only during the game, but also during the debriefing. The use of a virtual platform also eased the demand for physical space for huge class sizes. Medical students and nursing students are taught about emergency and urgent care at different years of their education. When they played Virtual ER, they were immersed in an ER environment that allowed them to perform triage, assess patients, formulate differential diagnoses, request investigations, and offer treatments. They were also able to recognize clinical deterioration and initiate resuscitation. During the debriefing, the teachers facilitated discussion with the students about managing patients and their shared experiences in handling patients in the ER. The debriefing provided opportunities for students to reflect, critique, and correct what they learned to further enhance their clinical competencies [34].

Another innovative aspect of this study was the assessment of the students' learning styles, which helped to determine whether there were any mismatches between teaching modes and learning styles. We found that the effect of Virtual ER on teamwork was much stronger in students who had more activist and pragmatist personality characteristics. In other words, they were more engaged in the game and gained more benefits in regards to teamwork enhancement from the game. Activists prefer learning by experience and taking direct action, while pragmatists like to experiment, see the relevance of their work, and adopt problem-solving approaches during learning [16]. Our game presented an immersive scenario in which a virtual patient was admitted to an ER with chest pain. The students discussed the case and chose the best investigation methods and treatments to save the patient's life. The problem-based learning structure of the game thus closely matched the preferences of activist and pragmatist learners. This also explains the positive primary outcome of this study: the game significantly enhanced teamwork. However, our results failed to support hypothesis 2, which was that students who were activists or pragmatists would have high game performance scores. This finding is in line with other studies, which have shown a lack of a direct relationship between learning experience and academic performance in medical students [35-37]. It is possible that learning style can

XSL•FO
RenderX

engage students in learning in serious games; however, it might not be a predictor of academic success. Criticism of the idea that adapting to individual learning styles can provide better academic performance can be found as early as 2004 [38]. In Virtual ER, students who tend to construct their own knowledge might have relied heavily on the debriefing, especially the opportunity for reflection, critique, and correction. Therefore, the game elements of Virtual ER and its design may not be adaptive enough to facilitate student learning. Alternative approaches to adapting serious games to cater to different students have recently been introduced, such as Felder-Silverman's model [39], which combined adaptive game-element designs for collaboration, gamification, and content interactions [40]. Although that study was designed for individual students interacting with the system, it would be a worthwhile approach to the further expansion of IPE among medical and nursing students in the future.

We found that learning styles were associated with teamwork enhancement but not clinical competency. This implies that individual students with different learning styles could find their own way to a more adaptive approach to learning. The students were encouraged to discuss the best clinical decisions for the virtual patient during class, and this discussion may have helped the students pick the correct answers and increase their performance. This may be beneficial, because as medical and nursing educators, we should provide a positive learning environment to facilitate individual students to develop their own learning approaches to optimize their academic performance. Nevertheless, in this study, students with prior advanced cardiovascular life support training showed a decrease in scores for teamwork attitude between baseline and after using the game. It is unclear whether students equipped with more advanced clinical knowledge tended to have more confidence and to work more on their own and therefore did not value the teamwork enhancement that Virtual ER made possible. This finding is consistent with a past study that showed that students had increased confidence in their own knowledge after basic life support training [41]. However, in the 21st century, in addition to academic and literacy skills, interprofessional collaborative practice is essential in the health care system. IPE is an important, innovative strategy for mitigating the global health workforce crisis, which has been reported by the World Health Organization [2]. Teamwork and collaboration skills cannot be built without engaging in practice and attitudinal changes. Therefore, educators should be creative in designing innovative pedagogical methods to match students' learning styles and engage them in learning essential "soft" skills, including teamwork, collaboration, communication, and leadership, for their personal and professional mastery. Serious games have potential for developing such attributes. Our findings demonstrate that our serious game positively increased teamwork attitudes and that this self-developed game should be a useful basis for further prospective expansion and enhancement. In the future, we would like to add learning goals on teamwork training. Furthermore, it will be possible to develop Virtual ER into an online platform, enabling IPE delivery regionally or globally.

Study Limitations

A few limitations weaken the generalizability of our results. First, this was a cohort study of 62 students without a comparison group, and the significant positive outcome may be attributable to the content of the game scenario, rather than to serious gaming in general. Therefore, a randomized controlled trial should be conducted in the future. We also did not examine the long-term effect of the serious game on teamwork enhancement, as our study was only focused on a single intervention. Future studies are needed to explore changes in teamwork attitude over a longer period, especially after graduation. Additionally, emergency medicine education is spread out in our medical curriculum, and the nursing students also had experience from part-time jobs as assistants in the health care system. Therefore, it is possible that some students may have already been exposed to some of the clinical conditions in the game, while others had not been. In this study, the majority of participants were medical students, and the effects of Virtual ER may have been unique in these students. Although we examined whether previous training in basic life support or advanced cardiovascular life support was associated with the study outcomes, it was still unclear whether a priori clinical experience affected some of the potential learning effects, nor was it clear how strong this effect might have been. In a future study with a larger sample size, we hope to perform a subgroup analysis to understand this relationship. Lastly, a future study could use a qualitative design to examine student satisfaction with the game to further improve it.

Conclusion

This study reports on the development of Virtual ER and evaluates its effect on teamwork enhancement in IPE and on clinical competency in handling patients with cardiovascular problems. Our findings demonstrate that this serious game had a positive effect on teamwork enhancement, in particular for students who had activist and pragmatist learning styles. We suggest that serious online games are a potential tool for the development of IPE and the fostering of 21st century life skills. Owing to its use of a virtual platform, this game might also be an innovative way to deliver IPE locally and globally.

Acknowledgments

The study was supported by a teaching and development grant (Project No. 750) from the University Grant Committee, Hong Kong, China. We are very grateful for the support of Ms Emily Leung, our graphic designer and game developer.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Kilner E, Sheppard LA. The role of teamwork and communication in the emergency department: a systematic review. Int Emerg Nurs 2010 Jul;18(3):127-137. [doi: 10.1016/j.ienj.2009.05.006] [Medline: 20542238]
2. Framework for action on interprofessional education and collaborative practice. World Health Organization. URL: https://www.who.int/publications/i/item/framework-for-action-on-interprofessional-education-collaborative-practice [accessed 2022-06-21]
3. Dow A, Thibault G. Interprofessional Education - A Foundation for a New Approach to Health Care. N Engl J Med 2017 Aug 31;377(9):803-805. [doi: 10.1056/NEJMp1705665] [Medline: 28854090]
4. Curran VR, Sharpe D, Flynn K, Button P. A longitudinal study of the effect of an interprofessional education curriculum on student satisfaction and attitudes towards interprofessional teamwork and education. J Interprof Care 2010 Jan 10;24(1):41-52. [doi: 10.3109/13561820903011927] [Medline: 19705318]
5. Wai AK, Lam VS, Ng ZL, Pang MT, Tsang VW, Lee JJ, et al. Exploring the role of simulation to foster interprofessional teamwork among medical and nursing students: A mixed-method pilot investigation in Hong Kong. J Interprof Care 2021 Dec 08;35(6):890-898. [doi: 10.1080/13561820.2020.1831451] [Medline: 33290116]
6. Wong A, Gang M, Szyld D, Mahoney H. Making an "Attitude Adjustment": Using a Simulation-Enhanced Interprofessional Education Strategy to Improve Attitudes Toward Teamwork and Communication. Simul Healthc 2016 Apr;11(2):117-125. [doi: 10.1097/SIH.0000000000000133] [Medline: 27043097]
7. Khalili H. Online interprofessional education during and post the COVID-19 pandemic: a commentary. J Interprof Care 2020 Jul 16;34(5):687-690. [doi: 10.1080/13561820.2020.1792424] [Medline: 32674721]
8. Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
9. Bedwell WL, Pavlas D, Heyne K, Lazzara EH, Salas E. Toward a Taxonomy Linking Game Attributes to Learning. Simul Gaming 2012 May 14;43(6):729-760. [doi: 10.1177/1046878112439444]
10. Dichev C, Dicheva D. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. Int J Educ Technol High Educ 2017 Feb 20;14(1):1-36 [FREE Full text] [doi: 10.1186/s41239-017-0042-5]
11. Landers RN. Developing a Theory of Gamified Learning: Linking Serious Games and Gamification of Learning. Simul Gaming 2015 Jan 15;45(6):752-768. [doi: 10.1177/1046878114563660]
12. Marks S, Windsor JA, Wünsche B. Design and evaluation of a medical teamwork training simulator using consumer-level equipment. Stud Health Technol Inform 2012;173:273-279. [doi: 10.3233/978-1-61499-022-2-273] [Medline: 22357001]
13. Steadman RH, Huang YM, Iseli MR, Lee JJ, Tillou A, Rudolph MD, et al. Screen-Based Simulation for Training and Automated Assessment of Teamwork Skills: Comparing 2 Modes With Different Interactivity. Simul Healthc 2021 Oct 01;16(5):318-326. [doi: 10.1097/SIH.0000000000000510] [Medline: 33086370]
14. Wilbur L. Interprofessional education and collaboration: a call to action for emergency medicine. Acad Emerg Med 2014 Jul 13;21(7):833-834 [FREE Full text] [doi: 10.1111/acem.12404] [Medline: 25125274]
15. Avrech Bar M, Katz Leurer M, Warshawski S, Itzhaki M. The role of personal resilience and personality traits of healthcare students on their attitudes towards interprofessional collaboration. Nurse Educ Today 2018 Feb;61:36-42. [doi: 10.1016/j.nedt.2017.11.005] [Medline: 29169067]
16. Honey P, Mumford A. The learning styles helper's guide. Maidenhead, UK: Peter Honey Publications; 2000.
17. Brannan JD, White A, Long J. Learning Styles: Impact on Knowledge and Confidence in Nursing Students in Simulation and Classroom. Int J Nurs Educ Scholarsh 2016 Aug 24;13(1):63-73. [doi: 10.1515/ijnes-2015-0052] [Medline: 27564700]
18. Fan K, Xiao P, Su C. The Effects of Learning Styles and Meaningful Learning on the Learning Achievement of Gamification Health Education Curriculum. Eurasia J Math Sci Technol 2015;11(5):1211-1229 [FREE Full text] [doi: 10.12973/eurasia.2015.1413a]
19. Morris SYJ. VISOLE: A constructivist pedagogical approach to game-based learning. In: Chi-Yin S, Harrison Hao Y, editors. Collective Intelligence and E-Learning 2.0: Implications of Web-Based Communities and Networking. Hershey, PA: IGI Global; 2010:185-206.
20. Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simul Gaming 2016 Aug 18;33(4):441-467. [doi: 10.1177/1046878102238607]
21. Beristain-Colorado MDP, Ambros-Antemate JF, Vargas-Treviño M, Gutiérrez-Gutiérrez J, Moreno-Rodriguez A, Hernández-Cruz PA, et al. Standardizing the Development of Serious Games for Physical Rehabilitation: Conceptual Framework Proposal. JMIR Serious Games 2021 Jun 24;9(2):e25854 [FREE Full text] [doi: 10.2196/25854] [Medline: 34185003]
22. Ak O. A Game Scale to Evaluate Educational Computer Games. Procedia - Social and Behavioral Sciences 2012;46:2477-2481 [FREE Full text] [doi: 10.1016/j.sbspro.2012.05.506]
23. Bellotti F, Kapralos B, Lee K, Moreno-Ger P. User Assessment in Serious Games and Technology-Enhanced Learning. Advances in Human-Computer Interaction 2013;2013:1-2 [FREE Full text] [doi: 10.1155/2013/120791]
24. Cooney A, Darcy E. 'It was fun': Exploring the pedagogical value of collaborative educational games. J Univ Teach Learn 2020 Jul 01;17(3):37-52 [FREE Full text] [doi: 10.53761/1.17.3.4]

XSL•FO
RenderX

25. Rahimi S, Shute V, Kuba R, Dai C, Yang X, Smith G, et al. The use and effects of incentive systems on learning and performance in educational games. Comput Educ 2021 May;165:104135. [doi: 10.1016/j.compedu.2021.104135]
26. Helmreich R. Reinforcing and evaluating crew resource management: Evaluator/LOS instructor reference manual. Austin, TX: NASA, University of Texas at Austin; 1990.
27. Chan CK, So H, Ng W, Chan P, Ma W, Chan K, et al. Does classroom-based crew resource management training have an effect on attitudes between doctors and nurses? Int J Med Educ 2016 Apr 09;7:109-114 [FREE Full text] [doi: 10.5116/ijme.56f5.6804] [Medline: 27060788]
28. Maric M, Penger S, Todorovic I, Djurica N, Pintar R. Differences in Learning Styles: A comparison of Slovenian Universities. Procedia Soc Behav Sci 2015 Jul;197:175-183. [doi: 10.1016/j.sbspro.2015.07.079]
29. Magezi DA. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). Front Psychol 2015;6:2 [FREE Full text] [doi: 10.3389/fpsyg.2015.00002] [Medline: 25657634]
30. Joseph S, Diack L. Playing interprofessional games: reflections on using the Interprofessional Education Game (iPEG). J Interprof Care 2015 May 07;29(3):260-262. [doi: 10.3109/13561820.2014.942839] [Medline: 25101519]
31. Chan A, Fung K, Orchard C. The circles of care game © - using gaming to teach interprofessional teamwork in clerkship. J Interprof Care 2020 Aug 23;34(1):133-136. [doi: 10.1080/13561820.2019.1639644] [Medline: 31438738]
32. Schmuck H, Arvin MK. Use of a board game format to promote interprofessional learning. J Interprofessional Educ Pract 2018 Sep;12:51-56. [doi: 10.1016/j.xjep.2018.05.007]
33. Friedrich C, Teaford H, Taubenheim A, Sick B. Interprofessional Health Care Escape Room for Advanced Learners. J Nurs Educ 2020 Jan 01;59(1):46-50. [doi: 10.3928/01484834-20191223-11] [Medline: 31945176]
34. Dreifuerst K. The essentials of debriefing in simulation learning: a concept analysis. Nurs Educ Perspect 2009;30(2):109-114. [Medline: 19476076]
35. Hamza M, Inam-Ul-Haq, Hamid S, Nadir M, Mehmood N. Effect of moderate learning style-teaching mode mismatch on academic performance among 2 year medical students in Pakistan. Indian J Psychiatry 2018;60(1):109-113 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_194_17] [Medline: 29736072]
36. Gurpinar E, Alimoglu MK, Mamakli S, Aktekin M. Can learning style predict student satisfaction with different instruction methods and academic achievement in medical education? Adv Physiol Educ 2010 Dec;34(4):192-196 [FREE Full text] [doi: 10.1152/advan.00075.2010] [Medline: 21098386]
37. Feeley A, Biggerstaff DL. Exam Success at Undergraduate and Graduate-Entry Medical Schools: Is Learning Style or Learning Approach More Important? A Critical Review Exploring Links Between Academic Success, Learning Styles, and Learning Approaches Among School-Leaver Entry ("Traditional") and Graduate-Entry ("Nontraditional") Medical Students. Teach Learn Med 2015 Jul 09;27(3):237-244. [doi: 10.1080/10401334.2015.1046734] [Medline: 26158325]
38. Price L. Individual Differences in Learning: Cognitive control, cognitive style, and learning style. Educ Psychol 2004 Sep;24(5):681-698. [doi: 10.1080/0144341042000262971]
39. Felder RM, Spurlin J. Applications, reliability and validity of the Index of Learning Styles. Int J Eng Educ 2005;21(1):103-112. [doi: 10.1037/t43782-000]
40. Hassan MA, Habiba U, Majeed F, Shoaib M. Adaptive gamification in e-learning based on students' learning styles. Interact Learn Environ 2019 Mar 17;29(4):545-565. [doi: 10.1080/10494820.2019.1588745]
41. Abelsson A, Odestrand P, Nygårdh A. To strengthen self-confidence as a step in improving prehospital youth laymen basic life support. BMC Emerg Med 2020 Jan 30;20(1):8 [FREE Full text] [doi: 10.1186/s12873-020-0304-8] [Medline: 32000691]

Abbreviations

ER: emergency room
HFAS: Human Factors Attitude Survey
IPE: interprofessional education
IPO: input-process-output
LMM: linear mixed effect modeling
NKDA: no known drug allergies

XSL•FO
RenderX

Edited by N Zary; submitted 29.11.21; peer-reviewed by B Wünsche, A Bustillo, M Graber; comments to author 03.02.22; revised version received 18.04.22; accepted 27.05.22; published 14.07.22.

Please cite as:
Wong JYH, Ko J, Nam S, Kwok T, Lam S, Cheuk J, Chan M, Lam V, Wong GTC, Ng ZLH, Wai AKC
Virtual ER, a Serious Game for Interprofessional Education to Enhance Teamwork in Medical and Nursing Undergraduates: Development and Evaluation Study
JMIR Serious Games 2022;10(3):e35269
URL: https://games.jmir.org/2022/3/e35269
doi:10.2196/35269
PMID:35834309

©Janet Yuen-Ha Wong, Joanna Ko, Sujin Nam, Tyrone Kwok, Sheila Lam, Jessica Cheuk, Maggie Chan, Veronica Lam, Gordon T C Wong, Zoe L H Ng, Abraham Ka-Chung Wai. Originally published in JMIR Serious Games (https://games.jmir.org), 14.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Augmented Reality–Based Guide for Mechanical Ventilator Setup: Prospective Randomized Pilot Trial

Sejin Heo[1,2], MD; Suhyeon Moon[3], MS; Minha Kim[1], MD; Minsu Park[4], PhD; Won Chul Cha[1,2], MD, PhD; Meong Hi Son[2,5], MD

[1]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[2]Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

[3]Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

[4]Department of Information and Statistics, Chungnam National University, Daejeon, Republic of Korea

[5]Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Corresponding Author:
Meong Hi Son, MD
Department of Pediatrics
Samsung Medical Center
Sungkyunkwan University School of Medicine
81, Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 34103539
Email: meonghi.son@samsung.com

Abstract

Background: Recently, the demand for mechanical ventilation (MV) has increased with the COVID-19 pandemic; however, the conventional approaches to MV training are resource intensive and require on-site training. Consequently, the need for independent learning platforms with remote assistance in institutions without resources has surged.

Objective: This study aimed to determine the feasibility and effectiveness of an augmented reality (AR)–based self-learning platform for novices to set up a ventilator without on-site assistance.

Methods: This prospective randomized controlled pilot study was conducted at Samsung Medical Center, Korea, from January to February 2022. Nurses with no prior experience of MV or AR were enrolled. We randomized the participants into 2 groups: manual and AR groups. Participants in the manual group used a printed manual and made a phone call for assistance, whereas participants in the AR group were guided by AR-based instructions and requested assistance with the head-mounted display. We compared the overall score of the procedure, required level of assistance, and user experience between the groups.

Results: In total, 30 participants completed the entire procedure with or without remote assistance. Fewer participants requested assistance in the AR group compared to the manual group (7/15, 47.7% vs 14/15, 93.3%; *P*=.02). The number of steps that required assistance was also lower in the AR group compared to the manual group (n=13 vs n=33; *P*=.004). The AR group had a higher rating in predeveloped questions for confidence (median 3, IQR 2.50-4.00 vs median 2, IQR 2.00-3.00; *P*=.01), suitability of method (median 4, IQR 4.00-5.00 vs median 3, IQR 3.00-3.50; *P*=.01), and whether they intended to recommend AR systems to others (median 4, IQR 3.00-5.00 vs median 3, IQR 2.00-3.00; *P*=.002).

Conclusions: AR-based instructions to set up a mechanical ventilator were feasible for novices who had no prior experience with MV or AR. Additionally, participants in the AR group required less assistance compared with those in the manual group, resulting in higher confidence after training.

Trial Registration: ClinicalTrials.gov NCT05446896; https://beta.clinicaltrials.gov/study/NCT05446896

(JMIR Serious Games 2022;10(3):e38433) doi:10.2196/38433

XSL•FO
RenderX

KEYWORDS

augmented reality; mechanical ventilation education; medical education; critical care; medical training; virtual reality; medical education; virtual education; nurse; nursing education; nursing; health care professional; learning platform; digital learning; digital health

Introduction

Mechanical ventilation (MV) is a lifesaving treatment that reduces the difficulty of breathing in patients and reverses acute life-threatening respiratory failure [1]. During the COVID-19 pandemic, the incidence of acute respiratory failure increased, leading to an increase in the demand for not only physical resources, such as ventilators and intensive care unit (ICU) beds, but also the ability to provide MV care expertise [2-4]. Effective and continuous MV education is important because adequate MV support improves clinical outcomes [5-8]. Regarding MV education, ICU nurses considered ventilator setup as an important topic; In ventilator setup, hands-on training is the most beneficial, suggesting that workshops or self-learning packages are not sufficient for novices to learn how to set up a ventilator [9]. However, conventional education usually focuses on theoretical knowledge (eg, prevention of infection and mode settings), and the type of hands-on training or bedside training that is required is human resource and time intensive, which limits educating several essential trainees [10-12].

Recently, augmented reality (AR) systems have been widely applied in medical education and training [13-16]. The AR system enables virtual objects to be overlaid onto a real-world environment by visualizing the physiological anatomy or enhancing the operator's view [17,18]. A few AR-guided medical procedure training regimes have been reported in the emergency department and intensive care environments [13,15,19-21]. They suggested that AR systems are effective in step-by-step procedures; however, the studies were limited to procedures lasting less than 10 minutes or to simple steps that did not reflect the usual complexity of procedures performed in the ICU or emergency department [22,23]. Additionally, limited research has been conducted on the independence or accuracy of the step-by-step procedures in AR systems [19].

In this study, we aimed to determine the effectiveness and feasibility of AR-based learning for novices to set up a ventilator by focusing on independently completing the procedures and assessing the degree of assistance required. Additionally, we evaluated the step characteristics in terms of the precision and assistance required.

Methods

This was a prospective randomized controlled pilot study conducted at Samsung Medical Center, Korea, from January to February 2022. We compared 2 modes of training, namely, the conventional method (via the printed manual) and the AR-based instructions. This study followed the CONSORT reporting guidelines (Multimedia Appendix 1).

Ethical Considerations

The research design was approved by the institutional review board of Samsung Medical Center (2021-12-112). Prior to inclusion in the study, all participants provided written informed consent.

Participants

We recruited nurses from the Samsung Medical Center who were interested in AR and ventilator education using a web-based hospital bulletin board. We enrolled nurses who had no prior experience with ventilator setup or AR systems, regardless of their work department or age. We excluded nurses who had already experienced setting up a ventilator or who had trouble wearing or using a head-mounted display (HMD). As this was a preliminary study, we were unable to determine the sample size. However, we referenced to past research on the step-by-step procedures with AR [19,24]. We set a target of 30 participants for recruitment.

Study Design

Using a lottery method, we randomly assigned the participants to 2 groups. One group (the manual group) used a printed manual to set up a ventilator and the other group (the AR group) used AR-based instructions through an HMD—HoloLens 2 (Microsoft Corporation). The participants in the AR group were provided with 15 minutes of learning and practice time with HoloLens 2. If they needed assistance, participants in the manual group made a phone call, and those in the AR group requested it remotely with HoloLens 2; subsequently, both groups were assisted by the same ICU nurse. In the AR group, the participants shared the same view as the nurse using Microsoft Dynamics 365 Remote Assist, which allowed the ICU nurse to guide the participants through voice commands and drawing marks on their view. Both groups were surveyed immediately after the task.

Instructions for Ventilator Setup

The instructions to set up the Servo-i mechanical ventilator (Maquet) were developed by researchers, including emergency physicians, pediatricians, and ICU nurses. The instructions detailed the entire process, from plugging in a socket to turning on the power, by performing initial ventilator mode setting with 35 steps. The AR instructions were developed as a step-by-step guide with the same text and images as in the printed manual, using the Microsoft Dynamics 365 Guide. The AR instructions were delivered using Microsoft HoloLens 2. The device allows users to go back and forth through the entire procedure by gazing at the screen when required. Some steps had a guide with a hologram of the 3D objects to indicate the location of the steps and direct the action of the connecting parts (Figure 1).

XSL•FO
RenderX

Figure 1. Scenes from the study: (A) trainee following the instructions with the printed manual; (B) trainee following the augmented reality (AR) instruction with head mounted display (HMD); (C) hologram instruction shown to trainee via HMD; (D) AR direction with hologram (3D object) overlaid on the mechanical ventilator.



Evaluation of Outcomes

The primary outcome is the overall score of the procedure, which is a 100-point scale converted from the original score. Participants scored 1 point for each step if they successfully finished the step within 5 minutes and obtained a maximum score of 35. The secondary outcome was the required level of assistance (ie, the number of steps and the number of participants who required assistance, assistance frequency, and assistance time). We also evaluated the user's experience with short questions on 3 themes: confidence, suitability, and whether they intended to recommend AR system to others [25,26]. All the participants were asked to respond to general questions on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The usability of the HMD in AR-based training was determined using previously validated system usability scale (SUS) standards [27,28].

Statistical Analysis

All continuous variables are described as mean (SD) and median (IQR), and categorical variables are described as numbers and percentages. For continuous variables, we used the Wilcoxon rank-sum test; for categorical values, we used the chi-square test or Fisher exact test. A proportion test was performed to compare the proportions between the two groups. For all statistical analyses, $P<.05$ was considered statistically significant. The statistical analysis was performed using the R software (version 4.1.2; R Foundation for Statistical Computing).

Results

Characteristics of the Participants

A total of 31 nurses with no prior experience in setting up ventilators were enrolled, with the exception of 1 participant who had an HMD-related technical issue. The remaining 30 nurses completed the entire procedure and the surveys afterward. Each participant was randomly assigned to either the manual or the AR group. The participants ranged in age from 24-53 years and came from a variety of departments: 4 participants from outpatient nursing, 9 participants from inpatient nursing, 11 participants from specialized nursing, and 6 participants from administrative and educational departments. There were no significant differences in sex, age, work experience, prior observation experience, or department of work between the two groups (Table 1).

XSL•FO
RenderX

Table 1. Demographics of the study participants (N=30).

Characteristics	Manual group (n=15)	Augmented reality group (n=15)	*P* value
Gender, n (%)			.48
Female	15 (100)	13 (86.7)	
Male	0 (0)	2 (13.3)	
Age (years), n (%)			>.99
20-30	9 (60)	8 (53.3)	
30-40	3 (20)	4 (26.7)	
≥40	3 (20)	3 (20)	
Age, median (Q1, Q3)	28 (26, 34.5)	28 (26.5, 35)	.84
Work (years), n (%)			.89
<5	7 (46.7)	7 (46.7)	
5-10	5 (33.3)	3 (20)	
>10	3 (20)	5 (33.3)	
Previous experience of observing, n (%)	0 (0)	1 (6.7)	>.99
Department of work, n (%)			.84
Outpatient nursing	3 (20)	1 (6.7)	
Inpatient nursing	4 (26.7)	5 (33.3)	
Specialized nursing[a]	5 (33.3)	6 (40)	
Administration and educational part	3 (20)	3 (20)	

[a]In operating room, emergency department, radiology department, intensive care unit, and imaging center.

Overall Performance

All 30 participants completed the entire procedure, with or without remote assistance. One participant in the AR group completed the procedure successfully without any assistance. Table 2 summarizes the overall results for the manual and AR groups. There was no significant difference in the overall score between the two groups, regardless of assistance. When only the steps without assistance were considered successfully passed, the median score was 88.57 (IQR 82.86-91.43) in the manual group and 91.43 (IQR 88.57-97.14; *P*=.10) in the AR group. However, if assisted steps were also considered as successful, the median score was 94.29 (IQR 91.43-94.29) in the manual group and 94.29 (IQR 92.86-97.14; *P*=.20) in the AR group.

The duration of the procedure between the two groups was not statistically significant. Without assistance, the median procedure time was 22.95 (IQR 19.37-24.69) minutes in the manual group and 23.95 (IQR 20.83-26.95; *P*=.60) minutes in the AR group. With assistance, the median procedure time was 25.32 (IQR 22.41-29.02) minutes in the manual group and 24.18 (IQR 22.37-28.41; *P*=.97) minutes in the AR group.

Multimedia Appendix 2 presents findings in terms of the step characteristics. We discovered that when following the directions of the ventilator or connecting and disconnecting materials such as a tube, circuit, or line, the manual group had a greater tendency to fail the steps or require assistance compared to the AR group. However, we did not identify any significant difference between the groups.

XSL•FO
RenderX

Table 2. Overall outcomes of augmented reality (AR)–based instructions (N=30).

Characteristics	Manual group (n=15)	AR group (n=15)	*P* value
Score, median (IQR)			
without assistance	88.57 (82.86-91.43)	91.43 (88.57-97.14)	.10
with assistance	94.29 (91.43-94.29)	94.29 (92.86-97.14)	.20
Procedure time (min), median (IQR)			
without assistance	22.95 (19.37-24.69)	23.95 (20.83-26.95)	.60
with assistance	25.32 (22.41-29.02)	24.18 (22.37-28.41)	.97
Assistance			
Steps that needed assistance, n	33	13	.004
Steps that needed assistance per participant, median (IQR)	2 (1-2)	0 (0-1.5)	.03
Participants who requested assistance, n (%)	14 (93)	7 (48)	.02
Assistance time (min), median (IQR)	1.53 (0.78-2.98)	0 (0-3.02)	.12

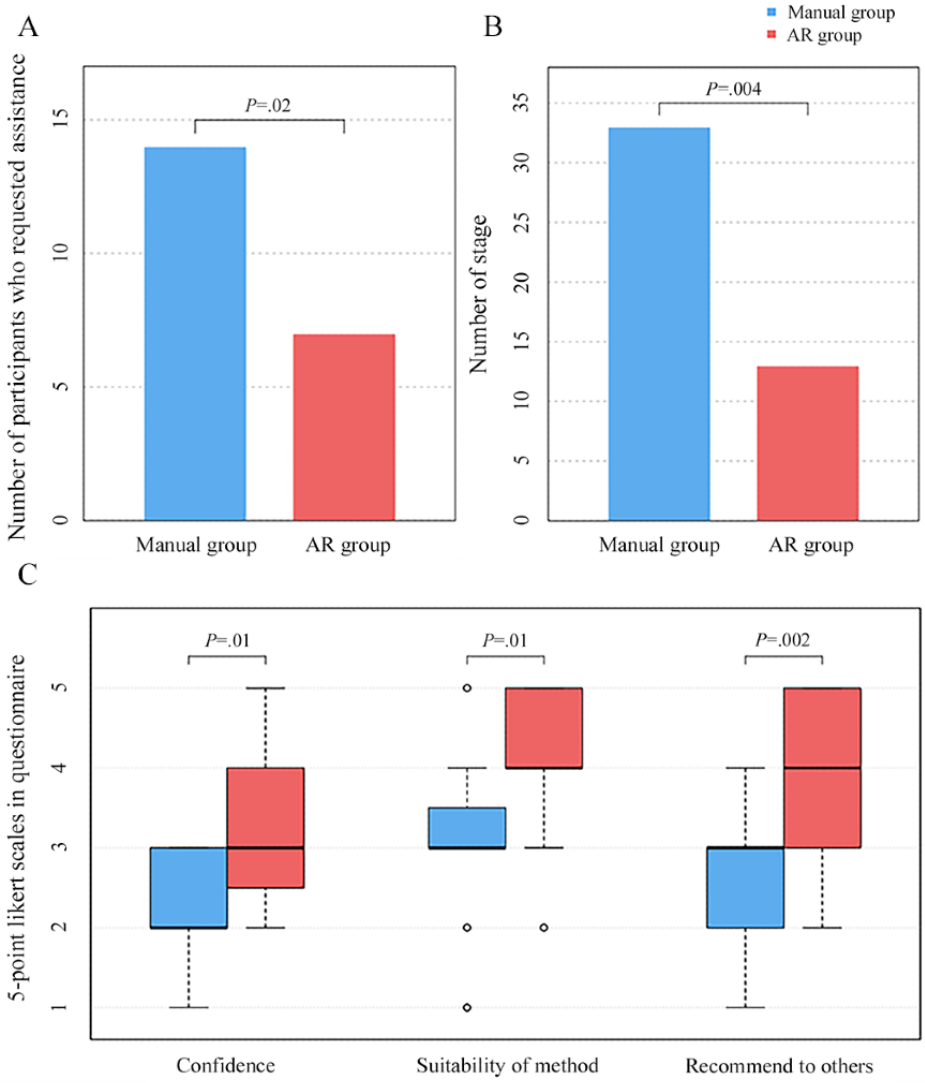
Need of Assistance

The manual group required considerably more assistance than the AR group. The median number of steps that required assistance per participant was greater in the manual group compared to the AR group (median 2, IQR 1-2 vs median 0, IQR 0-1.5; *P*=.03; Table 2).

The manual group had a greater proportion of participants who requested assistance compared to the AR group (14/15, 93.3% vs 7/15, 47.7%; *P*=.02; Figure 2A). Additionally, in the manual group, 33 requests for assistance were recorded, whereas only 13 requests were made in the AR group (*P*=.004; Figure 2B). There were no statistically significant differences in the time spent on assisting; the median was 1.53 (IQR 0.78-2.98) minutes for the manual group and 0 (IQR 0-3.02) minutes for the AR group (*P*=.12).

XSL•FO
RenderX

Figure 2. The effectiveness of augmented reality (AR)–based instruction for ventilator set up; (A) independent learning—the number of participants who requested assistance; (B) independent learning—number of stages that required assistance; (C) user experience.



Survey Outcomes

All the participants answered 3 general questions. Notably, only the AR group answered the SUS questions. Figure 2C shows the responses to the general questions on a 5-point Likert scale (from 1=strongly disagree to 5=strongly agree). The AR-based instructions received higher ratings for confidence (median 3, IQR 2.50-4.00 vs median 2, IQR 2.00-3.00; *P*=.01), suitability of method (median 4, IQR 4.00-5.00 vs median 3, IQR 3.00-3.50; *P*=.01), and whether they intended to recommend AR system to others (median 4, IQR 3.00-5.00 vs median 3, IQR 2.00-3.00; *P*=.002). The median of SUS score was 55 (IQR 47.5-67.5). Table 3 shows the details of each statement. Of all the statements, “well-integrated AR systems” received the best evaluation from the users, with the highest mean score and the lowest SD. Other statements such as “the simplicity of the system” (mean 3.7, SD 1.2), “ease of use” (mean 3.7, SD 1.1), and “technical assistance requirement” (mean 3.7, SD 1.1) obtained relatively high ratings.

XSL•FO
RenderX

Table 3. Evaluation of augmented reality (AR)–based instructions using the standardized system usability scale (SUS).

SUS questions	AR group (n=15), mean (SD)	AR group (n=15), median (IQR)
I think that I would like to use this system frequently.	3.5 (1.0)	4 (3.0-4.0)
I found the system unnecessarily complex.	3.7 (1.2)	4 (3.0-5.0)
I thought the system was easy to use.	3.7 (1.1)	4 (3.0-4.5)
I think that I would need the support of a technical person to be able to use this system.	3.7 (1.1)	4 (3.0-4.0)
I found that the various functions in this system were well integrated.	4.0 (0.6)	4 (3.0-5.0)
I thought there was too much inconsistency in this system.	1.9 (0.9)	2 (1.0-2.0)
I would imagine that most people would learn to use this system very quickly.	3.8 (1.0)	4 (3.0-5.0)
I found the system cumbersome to use.	2.4 (1.1)	2 (2.0-3.5)
I felt very confident using the system.	3.3 (1.0)	3 (2.5-4.0)
I needed to learn a lot of things before I could get going with this system.	3.3 (1.0)	4 (2.5-4.0)

Discussion

Principal Results

In this study, the participants had no prior experience with the ventilator setup or the HMD; additionally, all participants completed the entire procedure, from preparing materials to setting up the initial ventilator mode, prior to connecting to the patient. Moreover, the AR group was able to complete all the procedures following AR-based instructions in the planned design of the study, including a brief HMD practice and self-learning session. They required significantly less assistance compared to the manual group, and all assistance could be provided properly through a remote AR system. There were no technical issues or dropouts in either group.

Generally, hands-on training is required and beneficial when training trainees on complex procedures [29-31]. In addition, in a step-by-step procedure, failures in one step affect the subsequent steps, preventing the trainee from completing the process and requiring real-time guidance. However, experts in critical care cannot stay all day, and novices are required to use a systemic remote assistance when they face difficulties [32,33]. In addition, trainees do not have sufficient time to repeat the procedures, and when they practice alone, it is difficult to assist them in a proper manner.

AR-based training benefits both sides, as discussed above. The instructors are not required to be on-site, as the remote assistance software enables them to monitor and guide the trainee's view as well as draw and transfer data, voice, and so on. Additionally, without prior technical knowledge or experience with computer programming, developing instructions for a procedure using HoloLens 2 software was possible in 2-3 hours. From the trainee's perspective, they can learn frequently without visiting an education center or engaging in on-site instructions. Additionally, the desire to develop a contactless education platform has increased to protect health care workers and save on personal protection equipment [34].

The difference between the median number of assistance requests was not significant between the two groups; however, from a practical view, the difference in the workflow interruption between the two groups was more significant. A request for assistance resulted in procedure interruption by phone calls in the manual group and was difficult to support because they were held on to the phone. However, with an AR-based system, they could request help by speaking and connecting to the supervisor and be supported while continuing procedures.

Comparison With Prior Work

Our results provide new evidence for the feasibility of AR-based independent learning in complex advanced procedures with 35 steps lasting more than 20 minutes. In earlier research, studies also discovered that participants were more satisfied with AR-based instructions than with conventional instructions [35]. They were more confident and felt that they had received adequate training for the procedure, and they intended to recommend AR systems to others. It is important for continuous learning to attain competency [36]. We expect that strong confidence and user satisfaction would result in greater willingness and self-practice for learning to set up a ventilator independently.

Limitations

As a pilot study, there was no specific guideline regarding how to deal with technology issues, such as time for battery charging, overheating of the device without break time, and network instability. These issues were observed in a few cases but were solved without affecting the study; however, these issues will be addressed and planned in a larger-scale study.

Additionally, in the step-by-step procedures, the content of the errors is important; however, this was not addressed in this study. To extend AR-based training in other step-by-step advanced procedures and explore additional outcomes, considering the characteristics of steps and designing a training platform for suitable technology integration would be required.

Conclusions

AR-based instructions to set up mechanical ventilator were feasible for novices with no prior experience with MV and AR. In addition, participants using AR required less assistance, resulting in higher confidence after training.

XSL•FO
RenderX

Acknowledgments

We thank Minji Park and Kyung mi Min from the Research Institute for Future Medicine, Samsung Medical Center, Seoul, Korea for their assistance in conducting the study and for advising on the development of AR-based instructions.

Authors' Contributions

MHS contributed to the study's ideas and design and supervised the overall study progress. SH contributed to the enrollment of participants, prepared and conducted the study, analyzed the results, and wrote the draft. MSH and MP performed data analysis and interpreted the data. MK developed a short session of practice with AR systems and the survey questionnaire. WCC reviewed the study and the survey questionnaire. All the authors read and approved the final draft of the study.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered, as the authors did not believe registration to be necessary. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHealth checklist (V 1.6.1).
[PDF File (Adobe PDF File), 95 KB - games_v10i3e38433_app1.pdf]

Multimedia Appendix 2
Overall results of the study in a step-by-step manner.
[DOCX File , 21 KB - games_v10i3e38433_app2.docx]

References

1. Tobin MJ. Mechanical ventilation. N Engl J Med 1994 Apr 14;330(15):1056-1061. [doi: 10.1056/nejm199404143301507]
2. Wells CR, Fitzpatrick MC, Sah P, Shoukat A, Pandey A, El-Sayed AM, et al. Projecting the demand for ventilators at the peak of the COVID-19 outbreak in the USA. Lancet Infect Dis 2020 Oct;20(10):1123-1125. [doi: 10.1016/s1473-3099(20)30315-7]
3. Kwon AH, Slocum AH, Varelmann D, Nabzdyk CGS, MIT E-Vent Team. Rapidly scalable mechanical ventilator for the COVID-19 pandemic. Intensive Care Med 2020 Aug 25;46(8):1642-1644. [doi: 10.1007/s00134-020-06113-3] [Medline: 32588066]
4. Andellini M, De Santis S, Nocchi F, Bassanelli E, Pecchia L, Ritrovato M. Clinical needs and technical requirements for ventilators for COVID-19 treatment critical patients: an evidence-based comparison for adult and pediatric age. Health Technol 2020 Jul 30;10(6):1403-1411. [doi: 10.1007/s12553-020-00467-w]
5. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, et al. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000 Dec 04;342(18):1301-1308. [doi: 10.1056/NEJM200005043421801] [Medline: 10793162]
6. Adiyeke E, Koc BI, Coskun N, Bakan N, Sen E, Turan A, et al. The impact of a multidisciplinary education program for intensive care unit staff regarding ventilator care bundle on the frequency of ventilator-associated events. Dimens Crit Care Nurs 2021;40(4):210-216. [doi: 10.1097/dcc.0000000000000484]
7. Mireles-Cabodevila E, Dugar S. On the need for standard definitions and education to optimize patient-ventilator interactions. Respir Care 2017 Feb 20;62(2):248-249 [FREE Full text] [doi: 10.4187/respcare.05452] [Medline: 28108688]
8. Mogyoródi B, Dunai E, Gál J, Iványi Z. Ventilator-associated pneumonia and the importance of education of ICU nurses on prevention - preliminary results. Interv Med Appl Sci 2016 Dec;8(4):147-151 [FREE Full text] [doi: 10.1556/1646.8.2016.4.9] [Medline: 28180003]
9. Guilhermino MC, Inder KJ, Sundin D, Kuzmiuk L. Education of ICU nurses regarding invasive mechanical ventilation: findings from a cross-sectional survey. Aust Crit Care 2014 Aug;27(3):126-132. [doi: 10.1016/j.aucc.2013.10.064] [Medline: 24296161]
10. Guilhermino MC, Inder KJ, Sundin D. Education on invasive mechanical ventilation involving intensive care nurses: a systematic review. Nurs Crit Care 2018 Sep 26;23(5):245-255. [doi: 10.1111/nicc.12346] [Medline: 29582522]

XSL•FO
RenderX

11. Pedigo R. Ventilator management of adult patients in the emergency department. Emerg Med Pract 2020 Jul;22(7):1-20. [Medline: 32559026]
12. Haisley KR, Dunst CM, Kizziar P, Swanström LL. Impact of focused hands-on training course on practice adoption of advanced endoscopic techniques and per-oral endoscopic myotomy. J Laparoendosc Adv Surg Tech A 2022 Mar 01;32(3):251-255. [doi: 10.1089/lap.2021.0029] [Medline: 33794111]
13. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Dec;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
14. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 08;9(3):e29080 [FREE Full text] [doi: 10.2196/29080] [Medline: 34255668]
15. Munzer BW, Khan MM, Shipman B, Mahajan P. Augmented reality in emergency medicine: a scoping review. J Med Internet Res 2019 Apr 17;21(4):e12368 [FREE Full text] [doi: 10.2196/12368] [Medline: 30994463]
16. Williams MA, McVeigh J, Handa AI, Lee R. Augmented reality in surgical training: a systematic review. Postgrad Med J 2020 Sep 30;96(1139):537-542. [doi: 10.1136/postgradmedj-2020-137600] [Medline: 32229513]
17. Iqbal H, Tatti F, Rodriguez Y Baena F. Augmented reality in robotic assisted orthopaedic surgery: a pilot study. J Biomed Inform 2021 Aug;120:103841. [doi: 10.1016/j.jbi.2021.103841] [Medline: 34146717]
18. Cofano F, Di Perna G, Bozzaro M, Longo A, Marengo N, Zenga F, et al. Augmented reality in medical practice: from spine surgery to remote assistance. Front Surg 2021 Mar 30;8:657901 [FREE Full text] [doi: 10.3389/fsurg.2021.657901] [Medline: 33859995]
19. Wolf J, Wolfer V, Halbe M, Maisano F, Lohmeyer Q, Meboldt M. Comparing the effectiveness of augmented reality-based and conventional instructions during single ECMO cannulation training. Int J Comput Assist Radiol Surg 2021 Jul 23;16(7):1171-1180 [FREE Full text] [doi: 10.1007/s11548-021-02408-y] [Medline: 34023976]
20. Ingrassia PL, Mormando G, Giudici E, Strada F, Carfagna F, Lamberti F, et al. Augmented reality learning environment for basic life support and defibrillation training: usability study. J Med Internet Res 2020 May 12;22(5):e14910 [FREE Full text] [doi: 10.2196/14910] [Medline: 32396128]
21. Azimi E, Winkler A, Tucker E, Qian L, Doswell J, Navab N, et al. Can mixed-reality improve the training of medical procedures? Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:4065-4068. [doi: 10.1109/EMBC.2018.8513387] [Medline: 30441249]
22. Leary M, McGovern SK, Balian S, Abella BS, Blewer AL. A pilot study of CPR quality comparing an augmented reality application vs a standard audio-visual feedback manikin. Front Digit Health 2020 Feb 28;2:1 [FREE Full text] [doi: 10.3389/fdgth.2020.00001] [Medline: 34713015]
23. Rochlen LR, Levine R, Tait AR. First-person point-of-view-augmented reality for central line insertion training: a usability and feasibility study. Simul Healthc 2017 Feb;12(1):57-62 [FREE Full text] [doi: 10.1097/SIH.0000000000000185] [Medline: 27930431]
24. Azimi E, Winkler A, Tucker E, Qian L, Sharma M, Doswell J, et al. Evaluation of optical see-through head-mounted displays in training for critical care and trauma. 2018 Mar 22 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces; 18-22 March; Reutlingen, Germany. [doi: 10.1109/VR.2018.8446583]
25. Schrepp M, Hinderks A, Thomaschewski J. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). IJIMAI 2017;4(6):103. [doi: 10.9781/ijimai.2017.09.001]
26. Santos MEC, Chen A, Taketomi T, Yamamoto G, Miyazaki J, Kato H. Augmented reality learning experiences: survey of prototype design and evaluation. IEEE Trans Learning Technol 2014 Jan;7(1):38-56. [doi: 10.1109/tlt.2013.37]
27. Helin K, Kuula T, Vizzi C, Karjalainen J, Vovk A. User experience of augmented reality system for astronaut's manual work support. Front Robot AI 2018 Sep 12;5:106 [FREE Full text] [doi: 10.3389/frobt.2018.00106] [Medline: 33500985]
28. Wang C, Tsai N, Lu J, Wang MJ. Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks. Appl Ergon 2019 May;77:58-69. [doi: 10.1016/j.apergo.2019.01.007] [Medline: 30832779]
29. Sedlack RE, Petersen BT, Kolars JC. The impact of a hands-on ERCP workshop on clinical practice. Gastrointest Endosc 2005 Jan;61(1):67-71. [doi: 10.1016/s0016-5107(04)02456-3]
30. DiMusto PD, Chen H, Kent KC. Hands-on training in the operating room as a method of continuing education for surgeons in practice. JAMA Surg 2016 Sep 01;151(9):869-870. [doi: 10.1001/jamasurg.2016.1053] [Medline: 27462836]
31. Stovel RG, Gabarin N, Cavalcanti RB, Abrams H. Curricular needs for training telemedicine physicians: a scoping review. Med Teach 2020 Nov;42(11):1234-1242. [doi: 10.1080/0142159X.2020.1799959] [Medline: 32757675]
32. Khurrum M, Asmar S, Joseph B. Telemedicine in the ICU: innovation in the critical care process. J Intensive Care Med 2021 Dec;36(12):1377-1384. [doi: 10.1177/0885066620968518] [Medline: 33111599]
33. du Toit M, Malau-Aduli B, Vangaveti V, Sabesan S, Ray RA. Use of telehealth in the management of non-critical emergencies in rural or remote emergency departments: a systematic review. J Telemed Telecare 2017 Oct 05;25(1):3-16. [doi: 10.1177/1357633x17734239]
34. Greenberg JA, Schwarz E, Paige J, Dort J, Bachman S. At-home hands-on surgical training during COVID19: proof of concept using a virtual telementoring platform. Surg Endosc 2021 May 06;35(5):1963-1969 [FREE Full text] [doi: 10.1007/s00464-021-08470-6] [Medline: 33825008]

XSL•FO
RenderX

35. Siff L, Mehta N. An interactive holographic curriculum for urogynecologic surgery. Obstet Gynecol 2018 Oct;132 Suppl 1:27S-32S. [doi: 10.1097/AOG.0000000000002860] [Medline: 30247304]
36. Liang J, Wu S, Tsai C. Nurses' internet self-efficacy and attitudes toward web-based continuing learning. Nurse Educ Today 2011 Nov 31;31(8):768-773. [doi: 10.1016/j.nedt.2010.11.021] [Medline: 21176859]

Abbreviations

AR: augmented reality
HMD: head-mounted display
ICU: intensive care unit
MV: mechanical ventilation
SUS: system usability scale

Edited by N Zary; submitted 01.04.22; peer-reviewed by A Alkhaqani; comments to author 25.04.22; revised version received 01.05.22; accepted 12.06.22; published 22.07.22.

Please cite as:
Heo S, Moon S, Kim M, Park M, Cha WC, Son MH
An Augmented Reality–Based Guide for Mechanical Ventilator Setup: Prospective Randomized Pilot Trial
JMIR Serious Games 2022;10(3):e38433
URL: https://games.jmir.org/2022/3/e38433
doi:10.2196/38433
PMID:35867382

©Sejin Heo, Suhyeon Moon, Minha Kim, Minsu Park, Won Chul Cha, Meong Hi Son. Originally published in JMIR Serious Games (https://games.jmir.org), 22.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Points and the Delivery of Gameful Experiences in a Gamified Environment: Framework Development and Case Analysis

Sungjin Park[1*], PhD; Sangkyun Kim[1*], PhD

Graduate School of Business, Kyung Hee University, Seoul, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Sangkyun Kim, PhD
Graduate School of Business
Kyung Hee University
506, Orbis Hall
24, Kyungheedae-ro, Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 02 961 0448
Email: saviour@khu.ac.kr

Abstract

Background: Points represent one of the most widely used game mechanics in gamification. They have been used as a means to provide feedback to users. They visually show user performance and are used along with other game mechanics to produce synergy effects. However, using points without analyzing the application environment and targets adversely affects users.

Objective: This study aims to identify the problems that users encounter when points are applied improperly, to solve problems based on an analysis of previous studies and actual point use cases, and to develop a point design framework to deliver gameful experiences.

Methods: Three problems were identified by analyzing previous studies. The first problem is points that only accumulate. The second is points that emphasize a user's difference from other people. The third pertains to the reward distribution problem that occurs when points are used as rewards.

Results: We developed a framework by deriving 3 criteria for applying points. The first criterion is based on the passive acquisition approach and the active use approach. The second criterion is used to classify points as "high/low" and "many/few" types. The third criterion is the classification of personal reward points and group reward points based on segmentation of the reward criteria. We developed 8 types of points based on the derived point design framework.

Conclusions: We expect that some of the problems that users experience when using points can be solved. Furthermore, we expect that some of the problems that arise when points are used as rewards, such as pointsification and the overjustification effect, can be solved. By solving such problems, we suggest a direction that enables a gameful experience for point users and improves the core value delivery through gameful experiences. We also suggest a gameful experience delivery method in the context of the ongoing COVID-19 pandemic.

(JMIR Serious Games 2022;10(3):e35907) doi:10.2196/35907

KEYWORDS

point; design framework; gamification; gameful experience; pointsification; overjustification effect

Introduction

Background

Gamification is a technique that encourages participation and immersion by providing a gameful experience to users in nongame contexts, such as business management, marketing, health care, and education [1]. A gameful experience succeeds when users experience the feeling of playing a game [2] in what is normally a nongame context for them [1]. This is differentiated from a gaming experience, which refers to what an individual feels while playing an actual game; a gameful experience is a game-like experience in a nongame context [2]. A gameful experience enables users' enjoyment, absorption, creative thinking, absence of negative affect, activation, and dominance [2]. It is an important topic for millennials and

XSL•FO
RenderX

Generation Z (which we refer to collectively as "Generation MZ"), who differ from previous generations. They have benefited from technology, have been using smartphones from childhood, and have naturally accessed a variety of content through such devices. Games are a familiar type of content and are often accessed; consequently, people in these generations prefer gameful experiences [3,4].

Generation MZ has been forecast to become the largest consumer base from 2020 to 2035. Kristofer et al [5] analyzed the world's spending power and predicted the spending power of each generation from 2020 to 2035. According to this work, baby boomers and Generation X will show downward trends from 2020 to 2035, but Generation MZ is expected to continue to show an increasing trend [5].

Hence, many attempts have been made globally to serve Generation MZ, and gamification is one of those attempts. According to Park and Kim [6], education and training account for the largest portion of a total of 754 attempts to apply gamification worldwide, followed by human resource management, social issues, commercialization, and lifestyle.

As gamification began to draw attention, research began on systematizing development methods in order to deliver core values to users with gameful experiences. Gamification development consists of procedures that analyze constituent elements of a game, such as game mechanics [7], game rules, and sensory elements, that connect the player and the game physically and psychologically.

The oldest gamification development method is the point, badge, and leaderboard (PBL) system, which refers to game mechanics. The PBL system is the most widely used and easily applied method in gamification [8]. Park and Kim [6] showed that points are the most widely used game mechanic, and they are most commonly applied in gamification for health care [9] and in the field of education [10].

More factors should be considered than might be expected when applying gamification to deliver gameful experiences to users, including factors affecting the use of points, among other gamification elements. Environmental factors in applying points, user characteristics, and the significance and value of the points used as rewards should be considered. Before applying points, instructors should check if points are usable in a particular class. Additionally, it is necessary to consider whether missions and quests, which are required to give learners points, can be provided. Points are a means to stimulate intrinsic motivation and display the performance of a learner [11]. However, when points are applied without an appropriate purpose, they adversely affect users. Typical problems include "pointsification" and the "overjustification" effect. Pointsification is a phenomenon wherein the user does not perform the activity intended by the developer and acquires points without purpose, thus nullifying the purpose of applying gamification [11]. The overjustification effect is associated with motivation: when users act by their own will, it is because of inner motivation. In contrast, when users do not act on their own, reinforcement is used to increase the probability of the user performing the intended action [12]. By using points as reinforcement, users' external motivation is stimulated, and internalization of the external motivation is induced through continuous stimulation. The overjustification effect occurs when the internalization of external motivation is absent, and the internal motivation decreases because the user performs actions based only on the reinforcement [13,14].

This study analyzes problems with points reported in previous studies and develops a point-design framework to supplement the points system. Unlike the traditional methodology, the framework can be easily used by users, and it reduces the gap between the field and academic perspectives [15].

Literature Review

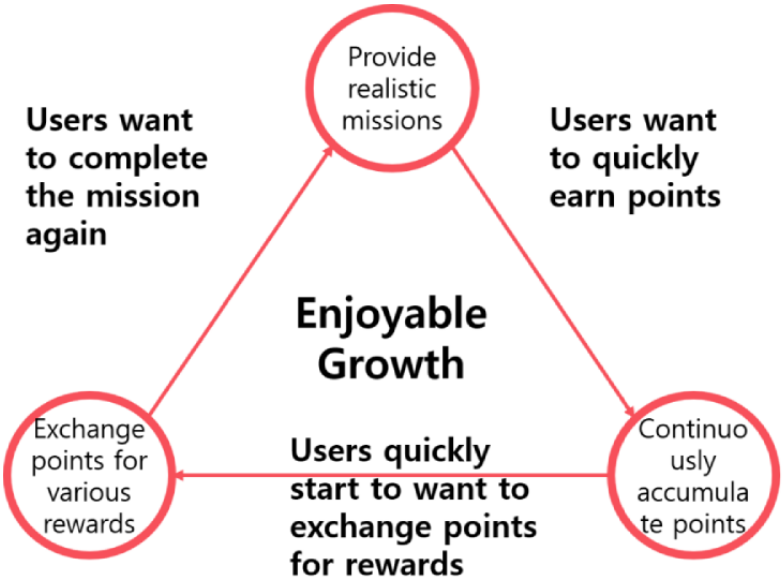
History and Effects of Points in Learning Environments

Points originated from a token economy that was used to induce changes in students' behavior at home or in the classroom. Tokens were issued to encourage behavior that helped learning activities, and the students exchanged the collected tokens for rewards that were helpful for studying. The token economy positively affected learners' self-reflection, improved their learning attitude, and stimulated participation in learning activities [16,17]. After 2011, when gamification was defined, the PBL system, which could add gameful experience to a token economy, was actively used, according to previous studies published between 2015 and 2020 [6,10,18]. Points provide learners with feedback by quantitatively showing feedback for learning activities. In a system described by Kim et al [14], learners received realistic missions from their instructors. The rewards that the learners received when they successfully performed these missions were points (Figure 1).

Points are accumulated when learners repeat a mission. The accumulated points represent the time and effort that the learners have invested in learning and show the role of stimulation on learners' motivation. Through this process, learners set goals for learning, compete with themselves and others in good faith, pursue self-achievement, and behave as expected by the instructor, thereby allowing behavior to be corrected through affordance [10].

XSL•FO
RenderX

Figure 1. The role of points in gamified systems.



Problems With Points in Learning Environments

If the point system is applied without considering cautionary factors, it may negatively affect learning. The use of points promotes learning activities, but the unconsidered use of points without an appropriate purpose has the effect of nullifying the gameful experience that gamification provides to users. We identified 3 problems with points reported in previous studies.

The first problem is the accumulation of points over time. In the token economy, tokens could be exchanged for rewards. Various methods were considered to prevent inner motivation from weakening due to this use of rewards. In a gamified environment, however, problems occur when points are used related to how the points that are obtained and accumulated by the learners are shown on the leaderboard. According to Kim et al [11], learners experience a sense of anticipation and achievement when they earn points. However, as points are accumulated, the sense of anticipation and achievement weakens, and the meaning of the points gradually fades. Consequently, learners no longer engage in point-earning activities. Furthermore, even if rewards that can be earned with points are provided, they can also have an adverse effect on learners, because new rewards that can stimulate more powerful external motivation must be continuously provided [11].

The second problem with points is that they can emphasize differences between people. When adopting the PBL system, many instructors display points that the learners earn on a leaderboard. In the system described by Park and Kim [6], the points display on the leaderboard showed the total points that were rewarded when a learner successfully completed a mission or quest set by the instructor. The learners checked and compared the displayed points, stimulating a competitive spirit. However, if the learners were exposed to this competitive environment for a long time, it resulted in academic stress, and the system turned into a simple competition that negatively affected learners [19]. Thus, points lose their feedback function, and learners work against the purpose of the gamification set by the instructor [20]. This is because the feedback provided by the points recorded on the leaderboard is the sum of all learning activities, not the individual activities of a learner. Learners grow through step-by-step feedback. Therefore, feedback that cannot provide specific details does not constitute effective feedback for learners [6].

The third problem is the problem of equally distributing rewards. In general, learning activities are divided into individual and team levels. Instructors set team-level quests to develop soft skills, such as collaboration, cooperation, and communication. A quest is established in such a way that all team members complete it together. If the team members perform the quest in cooperation, they earn points as rewards. Here, the team members qualitatively evaluate the activities of the participants in the team activity even though they cannot quantitatively evaluate them. If all team members are equally involved in the team activities, then each team member does not object to receiving the same number of points. However, if some of the team members do not participate diligently, then the other team members will feel that there is an equity problem. Based on this, the team members determine the fairness of the reward provided to each member. If the same reward is awarded to the "free riders," [21] who do not participate in the team activity and directly or indirectly impede the activities of other members, the other members perceive that the rewards are not fair and participate unenthusiastically in the learning activities. Further, if the rewards are the same for those who participate actively in the team activity but invest less time and effort compared to other learners because of personal ability issues, this also results in making them look like free riders to other learners.

Methods

Overview

In this section, we describe the development of a point-design framework to prevent the problems caused by points when instructors consider using gamification with the PBL system. The reason for developing such a framework is to ensure ease of use. A framework is a method of defining problems in engineering and logically explaining how to solve them [22].

XSL•FO
RenderX

The biggest advantage of a framework is that it reduces the gap between the field and academic perspectives. Therefore, we developed a framework that could address the problems of points and be applied in the field [15].

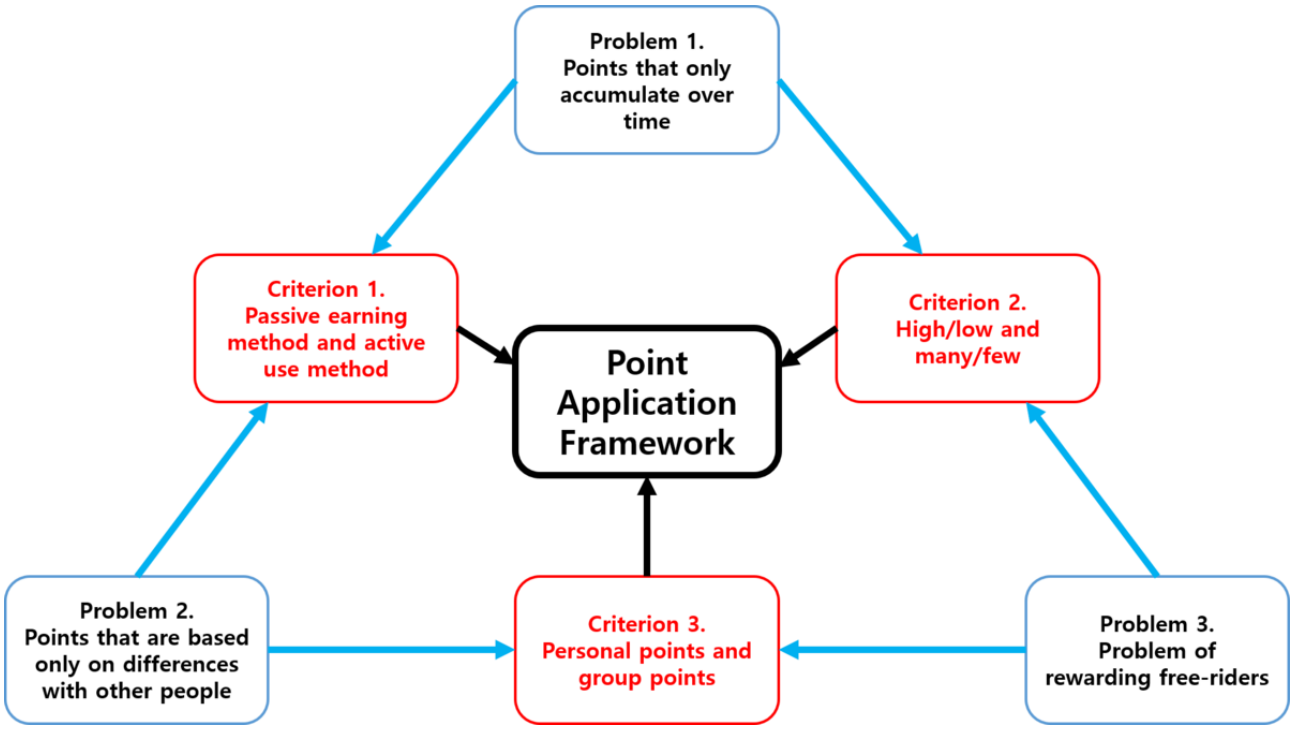
We designed a points application framework for gameful experiences in learning environments using 3 steps. The first step involved establishing 3 criteria to solve the problem of applying points in a learning environment, previously discussed in the literature review section.

The second step involved designing point-type criteria based on the criteria derived in the first step. The first step only derived a conceptual definition to solve the problem of points; however, it is also necessary to understand the users.

The final step involved defining the point types based on the results derived in the first and second steps. The point types were designed to be easily used by the instructor developing the gamified learning environment. To aid the understanding and use of the point types, we suggest examples of how to use the derived point types.

Our point design framework intends to solve the problems mentioned in the previous section. We derived 3 criteria for the application of points. As shown in Figure 2, the framework is focused on a design to solve 3 problems identified by previous studies.

Figure 2. Point application framework.



Criterion 1: Passive Earning Method and Active Use Method of Points

The instructor sets a mission or quest for the learner. Points are awarded differently depending on the difficulty of the mission or quest that the learner must perform. Points are meant to recognize the time and effort that the learner invested in completing the mission or quest. From the learner's perspective, they are an indicator of accomplishment, showing that the learner has completed a difficult job. The instructor should assign missions or quests to the learner repeatedly, so that the learner can master the knowledge. A problem arises in this process. The points earned by repeatedly completing missions or quests accumulate without being used. In the early stage of learning, the learner enjoys this, because the experience of earning points is new. However, with time, the joy derived from earning points diminishes.

The reason this phenomenon occurs is that points act as a reinforcement [12]. Meanwhile, the failure to provide better and better rewards leads to a degeneration of inner motivation [14]. From the standpoint of educators, it is practically impossible to periodically prepare better rewards. Reinforcements increase the probability that learners will perform the actions set by the educator just up to the point where the reward is given. Therefore, values other than reinforcement alone should be assigned to points.

The characteristics of points should be partially adjusted to use them as more than just reinforcements. The experience provided by points in the early learning stage is novel and stimulates a spirit of challenge. Therefore, the learners try to earn points. However, with time, they obtain points based on a sense of obligation because they are drawn to the familiarity of the points. Here, the experience, learning motivation, and sense of accomplishment are linked to familiarity [23]. The learners earn points because of this familiarity, and because the points become an "indicator" that displays their status on the leaderboard to other students, they no longer reflect the learners' willingness to learn, and the learners become passive.

XSL•FO
RenderX

Points should also change the learners' passive attitude to an active one. Conventional, passively earned points do not reflect learners' willingness. If the learners can use the points according to their own will, they will have a different experience. For example, let us assume that a learner has completed many missions and quests and collected a total of 30 points. The instructor can allow the learner to exchange the points for information that is helpful for learning, as a reward. In this case, the learner used the reward based on their own will, and this was an activity that was performed according to the rule agreed upon by everyone who participated in the learning activity.

By using points based on their own will, learners can improve the learning experience. In this way, learners will realize that there is a relationship between their existence and that of other people, and this will motivate the learners to actively participate in learning [24]. Furthermore, the active use of points stimulates the cohesion of the structural relationship of learners while interacting with the educational environment set up by the instructor. This promotes learning activities, and as a result, points act as a genuine reinforcement [14,24].

In the end, the passive earning method of points should be utilized for conventional feedback functions and triggering competition in good faith, while the active use method should be used to improve the learning experience of learners and solve the problems caused by accumulated points to let the learners become immersed in studying. When learners successfully complete missions or quests using both the passive earning method and the active use methods, their self-efficacy will be maximized. Learners will believe that the next mission or quest can be successfully completed, and they will also believe in the value of the reward received after completing a mission or quest in the group that they belong to [14]. Through this process, the learners' learning experience will improve, positively affecting their learning.

Criterion 2: Significance of Numbers in Points—High/Low and Many/Few

Points have the characteristic of being numbers. Therefore, when they are displayed on the leaderboard, or when users interpret their meaning, they reflect the characteristics of numbers. Numbers can express both quality, as simple numbers, and quantity, as a number of tokens, coupons, or coins. Qualitatively, if player A has a higher score than player B, then we acknowledge that player A is the one who did better. Furthermore, players with a lower score will be motivated in their activities and will set the goal of breaking the record of the player with the highest score [25]. These are points in the "high/low" sense. Meanwhile, points can also be used quantitatively, as in the expression that player A has earned "many points" while player B has earned "few points." Quantitative points are similar to tokens in the token economy system. In this system, if the intended activity of the educator is performed, tokens are received, and the earned tokens are used and exchanged for a reward [26]. For high/low points, scores can are displayed on a leaderboard, and users compare their scores to others'. Quantitative points allow a relative comparison of many/few points without a leaderboard, based on the quantity of points that a user possesses.

The high/low points displayed on the leaderboard provide intuitive feedback to the learner. Meanwhile, many/few points are related to operant conditioning. In a gamified learning environment, rewards are awarded differently depending on the activity that learners select. Here, points are used as a reinforcement to encourage learners to perform the activities intended by the educator. As time passes, the learners accumulate experience with points. The learners can use points for optional rewards by operant conditioning rather than as reinforcements. Based on operant conditioning, quantitative points encourage learners to actively participate in learning activities [27].

The concept of quantitative points is different from that of virtual currencies. A virtual currency is a type of game mechanic associated with the economic system of a game, and the educator must maintain fairness through currency balance [14]. It is also a kind of game mechanic used for compensation. However, quantitative points induce the action of exchanging points for rewards set by the educator using the concept of exchange. Therefore, with virtual currency, there is no price or benefit, and the exchanged rewards do not lead to the collapse of the learning balance between learners. Therefore, quantitative points and virtual currencies are different concepts.

Criterion 3: Personal and Group Points

Slavin [28] insists that systematically establishing a reward structure for learners is important because, according to their analysis, if an environment for sharing the learning activity process and results with other learners is established, a personal-level reward structure will increase learners' performance. Considering the social aspect of learners, however, a cooperative reward structure is needed. This is because learners are attracted to the soft skills required for the interaction between learners, including skills such as learning and building social rapport, rather than simply acquiring knowledge. Ultimately, the classroom atmosphere can be changed. Social activities among learners have positive effects on communication, listening ability, problem-solving skills, and learning motivation [29]. Therefore, educators should simultaneously consider both the personal level and the group level when they design learning activities.

No major problems occur with the personal points received after completing personal-level learning activities, because the learner is rewarded in proportion to the time and effort invested. However, the fairness of rewards is important in activities where groups of 2 or more people participate. Suppose student A performed a group activity as well as they could, but student B did not. If they both receive the same level of reward, student A will not participate properly in subsequent learning activities. In this process, student A loses the satisfaction and value of the reward, because their belief and expectation in the the fairness of the reward are damaged, and its significance is lost. Due to this damage to motivation, the possibility of lukewarm participation will increase [30]. Therefore, learning activities should be subdivided, and personal points and group points should be configured to determine the points awarded to the group and the points awarded to individuals, even for group activities.

XSL•FO
RenderX

Principle of Solving Problems in the Point Design Framework

The 3 criteria derived in this study solve the 3 problems based on the following principle: The passive point-earning method, in which points are only accumulated over time, is associated with the characteristic of numbers that they can be high or low, because the points are recorded on the leaderboard to deliver feedback. Based on how high or low the score is, the participation level in the learning activity can be understood, and the learning level of the learner can be understood. Meanwhile, if the active point-use method is added to the passive point-earning method, the problem arising from point accumulation disappears. By exchanging the points earned for a reward offered by the educator, learners perceive that they were given a new means or privilege to participate actively in learning activities. Therefore, if the passive earning method and the active use method are combined and applied to points, and the meaning of high/low and many/few points is used, a new experience can be delivered to learners. Furthermore, while classifying the size of the rewards at the personal and group levels, if additional missions or quests are set up so that personal-level rewards can also be received in a group learning activity and additional rewards are configured to be awarded differently depending on the participation level of individuals in the group activity, the free-rider problem can be partially solved. Additionally, the ability to make strategic choices can be granted to learners through various point types. When learners choose their own learning activities to earn points, they begin to understand the meaning of points. Here, if learners choose learning activities with point types related to the active use method, they will earn rewards that are helpful for the learning activity. For strengthening competence and competition in good faith, however, learners prefer learning activities in which points can be passively earned. As such, learners will strategically participate in learning, which is expected to improve their learning attitudes.

Results

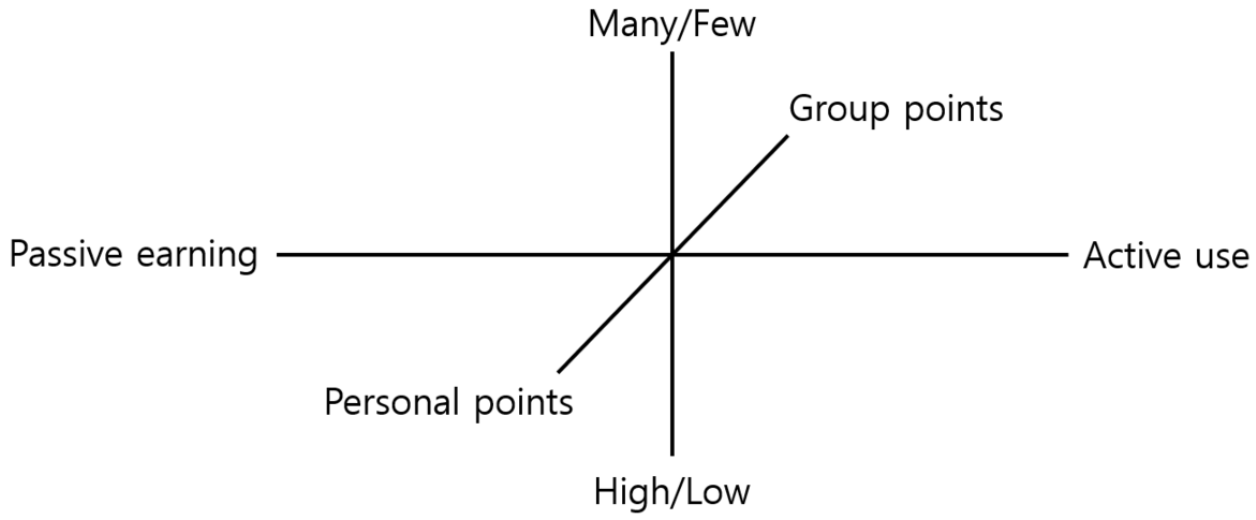
Point Types and Examination of Actual Applied Cases

We derived a point application framework based on the 3 criteria. Based on the framework, we derived 8 types of points. The point types were classified and derived based on the criteria shown in Figure 3.

The first point type is experience points (EXP). EXP represents the player's activity numerically. It is a game mechanic related to the user's level. It can be earned passively and compared as a many/few–type number. Furthermore, EXP is given as a personal reward. As players collect more EXP, their level goes up, and the time and effort that they invested can be estimated based on how high their level is. Figure 4 shows an example of EXP in Duolingo (Duolingo, Inc), a language-learning platform. This platform expresses experience points as "XP." If the specified amount of learning is completed, XP can be earned. The user cannot control their XP. Learning motivation is stimulated by achieving a higher level through XP.

The second point type is guild EXP. Guild is a word that originates from the guild system that craftsmen established to train apprentices in medieval Europe. Learners belonging to a guild expand their knowledge and skills by interacting with guild members [30]. Guild EXP is given as a group reward, unlike regular EXP. Guild EXP is also used with levels and uses a structure in which guild members work hard to accumulate guild EXP to increase the guild's level. As the guild's level increases, value-added effects increase. For example, the guild becomes envied by other players, who set a goal to join the guild. Figure 5 shows a clan leaderboard of CodeCombat (CodeCombat Inc), a platform for programming-education content. This platform uses the word "clan" instead of "guild," and the number of heroes corresponds to guild EXP. The number of heroes is derived from the missions and quests completed by learners belonging to the clan.

Figure 3. Point type criterion.



XSL•FO
RenderX

Figure 4. Experience point example (Duolingo).

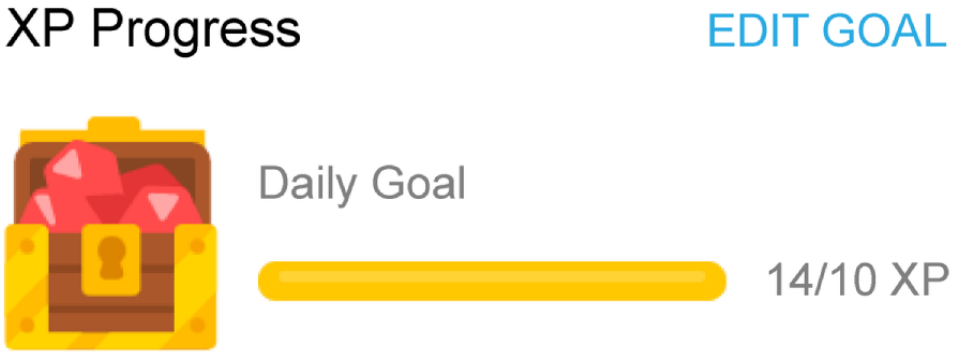
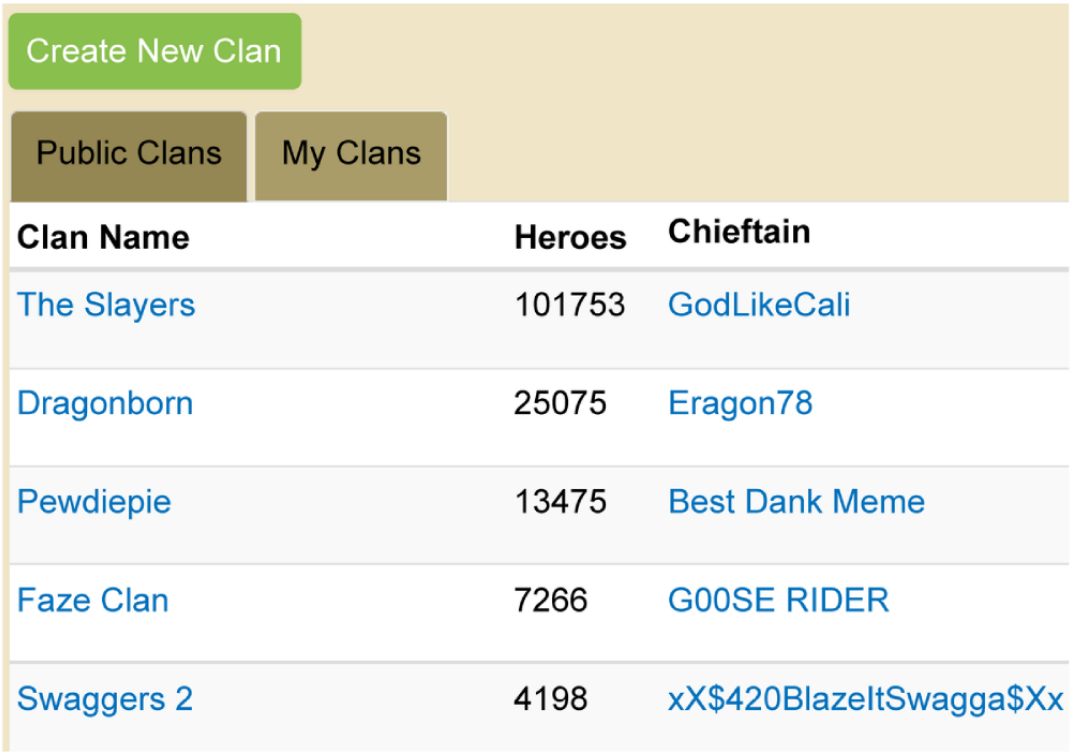


Figure 5. Guild experience point example (CodeCombat).

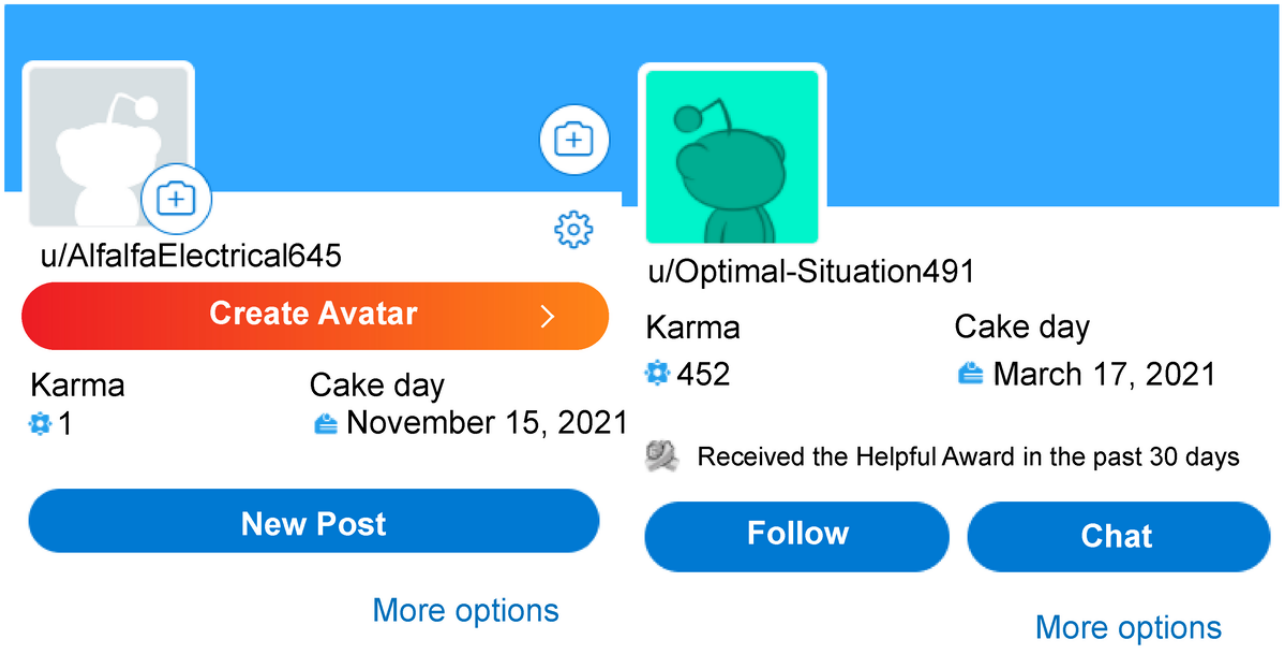


The third point type is ability points. Ability points can be earned passively, and the points can be compared as high/low–type points. Furthermore, they are personal reward points. Ability points visually show that knowledge or skill in the relevant area has increased as a result of actively performing learning activities when learning various subjects, such as English, mathematics, and leadership. For example, if student A studies a higher level of mathematics compared to other subjects, the instructor raises the logic level of student A to the next level. Based on this process, the learner's competence can be represented numerically by changing the learner's abilities (eg, wisdom, power, and logic) [31]; the abilities that the learner can obtain by studying are set to dexterity, intelligence, and discipline. The relevant ability points can be increased after completing actual missions and quests.

The fourth point type is karma points. Karma points can be earned passively and compared as high/low points, but they are group reward points. If some members in a team of 2 or more learners successfully complete a mission or quest, all members of the team receive points as a reward. Karma points can stimulate peer companionship, and using them as a feedback device among group members induces changes in the learning behavior of the beneficiary learners [32]. In the process, the learners will have a positive learning experience and feel fulfillment, pride, and satisfaction with the learning activity based on the bonds they form with other learners [14,33,34]. Figure 6 shows a user screen from Reddit (Reddit Inc), a social media service. When a user posts or shares something meaningful to other people, other users send karma points as a way of appreciating the user's activity. When users' karma points are high, it implicitly acknowledges that they have provided meaningful information to many people.

XSL•FO
RenderX

Figure 6. Karma point example on Reddit.



The fifth point type is exchange points. Those who possess points can actively use them, and the more points, the better. They can also be exchanged for something for oneself, such as virtual items, benefits, or privileges that are helpful for learning according to the reward system set by the educator. Exchange points differ from virtual currencies; point users use them based on their own thought and will. Figure 7 shows a virtual shop in CodeCombat. Players can earn blue crystals as points by performing missions or quests and exchanging the earned points for items that can make their avatar more powerful.

The sixth point type is group exchange points. Unlike exchange points, the entity that uses the points is not the individual, but the group, based on their collective opinion. The privileges, benefits, and virtual items exchanged using points should be beneficial to all members. This point type is commonly used in massively multiplayer role-playing games, and the group leader engages in discussions with the members to exchange the points for items or effects they need.

The seventh point type is skill points. Skill points are used when encouraging learners to obtain certain knowledge or skills. For example, suppose the learners wish to learn theory A, and the instructor deliberately sets up relevant missions and quests. Here, the instructor classifies the theory A into levels 1, 2, and 3. If learners complete a mission or quest related to learning theory A several times, they can attempt a higher level of mission or quest. If the learner has successfully completed the level 1 mission or quest, the learning level of theory A is raised by 1 point. If all 3 points have been earned by repeating this process, other people acknowledge that the learner understands theory A to a corresponding degree. Furthermore, when the skill points increase, privileges or benefits may be granted to help the learner with learning. For example, a learner who has achieved 2 skill points for theory A may help another learner when performing the level 1 mission or quest or use the privileges or benefits acquired to overcome a difficulty that they face. Figure 8 shows an example of skill points in Classcraft (Classcraft Studios Inc), a class management program. Skills exist in the form of a tree, and a user's points in a lower hierarchical skill must reach a certain level to use an upper hierarchical skill. If skill points increase, better benefits or privileges can be experienced.

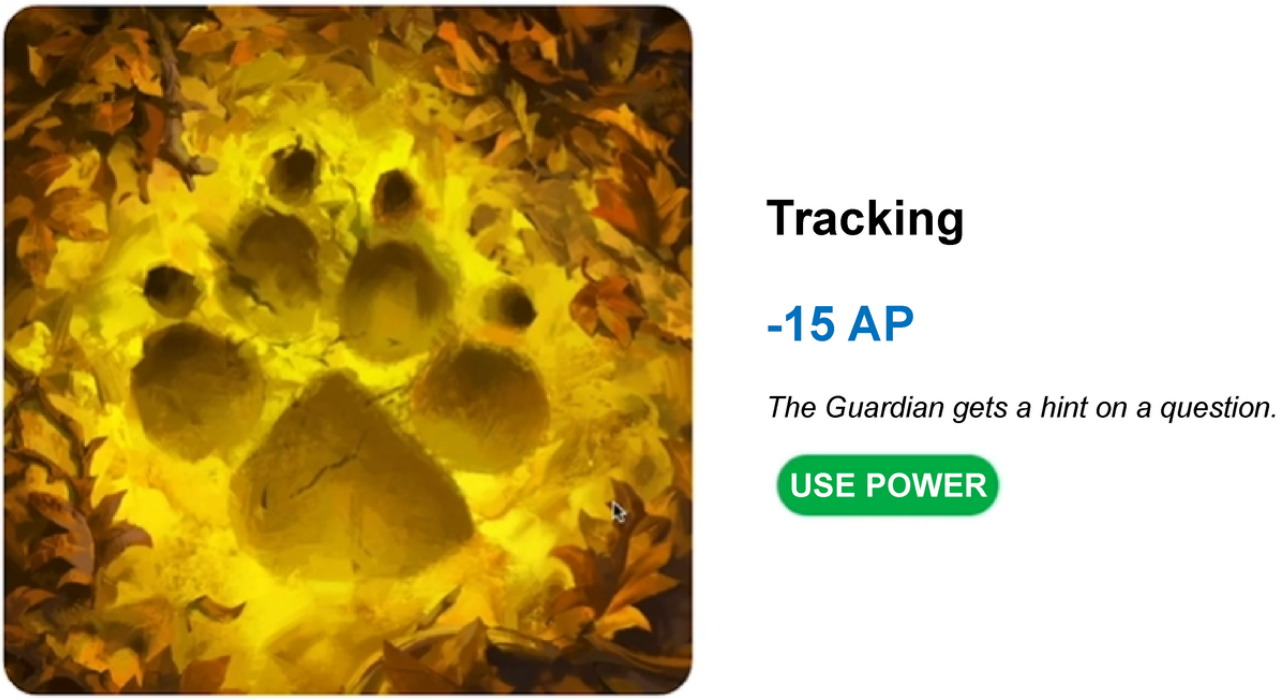
The eighth point type is peer review points, which are used when learners in teams of 2 or more evaluate each other's learning activity. Whereas karma points are emotionally linked to the members, peer review points are used when evaluating peers. The members of other teams cannot use them, and these points are used when evaluating the members of the same team based on the experience of interacting with them. Peer review points create bonds between members by letting them provide feedback to each other, and based on the bonds, learners build their own learning experiences. This process positively affects learning [35].

XSL•FO
RenderX

Figure 7. Exchange point example (CodeCombat).



Figure 8. Skill point example (Classcraft).



XSL•FO
RenderX

Discussion

Points are a gameful experience–delivering element that can be used in the largest proportion of the educational environment. The performance of learners can be visually displayed through points, and the displayed points can be used as the basis for learners to recognize each other's progress. Educators use rewards to set up goals and stimulate learning motivation so that the learners can study smoothly. The learners recognize the points set up by the instructor and participate in learning activities, whereby it is hoped that the learning attitude of the learners will improve. However, improper use of points can adversely affect the learners and the educator. For example, the improper use of points may cause learners to nullify gamification and participate in learning activities to earn points without a genuine purpose.

The aim of this study is to suggest guidelines, by developing a point design framework, for educators considering the application of points that will help them avoid mistakes. To this end, we explore the educational effects of points and the definition of the framework and have presented 3 problems arising due to points. To solve these 3 problems, we have proposed 3 design criteria and 8 types of points based on past studies and actual cases.

We recommend the following approaches to educators who are considering applying points and using the point application framework developed in this study. First, the reinforcement and reward characteristics of points must both be reflected. If points are used as simple reinforcements, the learners will only repeat the learning activity intended by the instructor. Consequently, there is a high possibility that the learners will not experience joy as time passes. For sustainable learning, therefore, the point application framework developed in this study should be used to design points that have both reinforcement and reward characteristics. Second, attention should be paid to pointsification. It is necessary to choose actions through which the learners earn points, but if the points have more value than the learning activity itself, the learners will try to earn points without a true purpose. Therefore, points should use a sophisticated design based on the point application framework. Otherwise, learners will feel lethargic and find problems regarding fairness, and if these problems are not resolved, the learners will give up on learning. Last, the point application framework should be used to prevent the overjustification effect. Even high inner motivation in a learner may become weakened if the learner is continuously exposed to external rewards. In this case, the learner may blame the rewards for their change in attitude [36]. The point application framework and mission or quest settings should be considered at the same time to prevent the overjustification effect. Points should not be used as simple rewards, and an environment that promotes the stimulation of learning motivation should be established based on operant conditioning [37].

There were limitations of this study that suggest future research directions. If a badge application framework and a leaderboard design framework are used, it is possible to develop a PBL system that promises educational effects while effectively delivering a gameful experience. However, there is a lack of studies on missions and quests related to PBL. Research on methods for setting up missions and quests for the effective education of learners is insufficient, and this aspect is not covered in this study. Further research is required to develop a methodology to set missions and quests or a framework that will facilitate the balanced education of learners.

Acknowledgments

This research was supported by a grant from the National Research Foundation of Korea and funded by the Korean Government Ministry of Science and Information and Communication Technologies (grant 2020R1A2B501001801). This research was also supported by the Korean Government Ministry of Science and Information and Communication Technologies under its Information Technology Research Center support program (grant IITP-2021-0-02051) supervised by the Institute for Information & Communications Technology Planning & Evaluation.

Conflicts of Interest

None declared.

References

1. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sept 28-30, 2011; New York, NY p. 9-15. [doi: 10.1145/2181037.2181040]
2. Eppmann R, Bekk M, Klein K. Gameful experience in gamification: construction and validation of a gameful experience scale [GAMEX]. J Interact Mark 2018 Aug;43:98-115. [doi: 10.1016/j.intmar.2018.03.002]
3. Park S, Min K, Kim S. Differences in learning motivation among bartle's player types and measures for the delivery of sustainable gameful experiences. Sustainability 2021 Aug 14;13(16):9121 [FREE Full text] [doi: 10.3390/su13169121]
4. Turner A. Generation Z: technology and social interest. J Individ Psychol 2015;71(2):103-113. [doi: 10.1353/jip.2015.0021]
5. Kristofer H, Katharina F, Martin H. How to harness the spending power of millennials: Move beyond the US. Brookings Institution. 2018. URL: https://tinyurl.com/yckxw94v [accessed 2022-08-31]

XSL•FO
RenderX

6. Park S, Kim S. Patterns among 754 gamification cases: content analysis for gamification development. JMIR Serious Games 2018 Nov 06;6(4):e11336 [FREE Full text] [doi: 10.2196/11336] [Medline: 30401666]
7. Sicart M. Defining game mechanics. Game Studies 2008;8(2):1-14 [FREE Full text]
8. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business. Upper Saddle River, NJ: Wharton Digital Press; 2012.
9. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
10. Zainuddin Z, Chu SKW, Shujahat M, Perera CJ. The impact of gamification on learning and instruction: A systematic review of empirical evidence. Educ Res Rev 2020 Jun;30:100326. [doi: 10.1016/j.edurev.2020.100326]
11. Kim S, Song K, Lockee B, Burton J. Legal and ethical issues. In: Gamification in learning and education. Advances in Game-Based Learning. Cham, Switzerland: Springer; 2018:109-116.
12. Gordan M. A Review of B. F. Skinner's 'Reinforcement Theory of Motivation'. Int J Res Educ Methodol 2014 Dec 30;5(3):680-688. [doi: 10.24297/IJREM.V5I3.3892]
13. Tang S, Hall VC. The overjustification effect: A meta-analysis. Appl Cognit Psychol 1995 Oct;9(5):365-404. [doi: 10.1002/acp.2350090502]
14. Kim S, Song K, Lockee B, Burton J. Theories for gamification in learning and education. In: Gamification in learning and education. Cham, Switzerland: Springer; 2018:39-47.
15. Leshem S, Trafford V. Overlooking the conceptual framework. Innov Educ Teach Int 2007 Feb;44(1):93-105. [doi: 10.1080/14703290601081407]
16. Zlomke K, Zlomke L. Token economy plus self-monitoring to reduce disruptive classroom behaviors. Behav Anal Today 2003;4(2):177-182. [doi: 10.1037/h0100117]
17. Nelson KG. Exploration of classroom participation in the presence of a token economy. J Instr Psychol 2010;37(1):49-56.
18. Dicheva D, Dichev C, Agre G, Angelova G. Gamification in education: A systematic mapping study. J Educ Techno Soc 2015;18(3):1-14.
19. Lätsch A. Do perceived helpfulness and competition in classroom contexts moderate or mediate the association between perceived stress and socio-emotional strengths and difficulties from early to middle adolescence? Learn Individ Differ 2017 Aug;58:31-40. [doi: 10.1016/j.lindif.2017.07.006]
20. van Roy R, Zaman B. Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In: Serious Games and Edutainment Applications. Cham, Switzerland: Springer; 2017:485-509.
21. Fobes C, Kaufman P. Critical pedagogy in the sociology classroom. Teach Sociol 2008 Jan 01;36(1):26-33. [doi: 10.1177/0092055x0803600104]
22. Johnson R, Deutsch P. How to design frameworks. Notes for OOPSLA. 1993. URL: https://www.researchgate.net/publication/243763805_How_to_Design_Frameworks [accessed 2021-11-30]
23. Krugman H, Hartley E. Passive learning from television. Public Opin Q 1970 Jan 1;34(2):184-190. [doi: 10.1086/267788]
24. Kim J. Where Technology Meets Design and Strategy: Design for Experience. Cham, Switzerland: Springer; 2015.
25. Furnham A, Boo HC. A literature review of the anchoring effect. The Journal of Socio-Economics 2011 Feb;40(1):35-42. [doi: 10.1016/j.socec.2010.10.008]
26. McLaughlin TF, Malaby J. Intrinsic reinforcers in a classroom token economy. J Appl Behav Anal 1972;5(3):263-270 [FREE Full text] [doi: 10.1901/jaba.1972.5-263] [Medline: 16795348]
27. Staddon JER, Cerutti DT. Operant conditioning. Annu Rev Psychol 2003 Feb;54(1):115-144 [FREE Full text] [doi: 10.1146/annurev.psych.54.101601.145124] [Medline: 12415075]
28. Slavin RE. Classroom reward structure: an analytical and practical review. Rev Educ Res 2016 Jun 30;47(4):633-650. [doi: 10.3102/00346543047004633]
29. Sun-Lin H, Chiou G. Effects of gamified comparison on sixth graders' algebra word problem solving and learning attitude. J Educ Techno Soc 2019 Jan;22(1):120-130 [FREE Full text]
30. Bui B, Porter B. The expectation-performance gap in accounting education: an exploratory study. Account Educ 2010 Apr;19(1-2):23-50. [doi: 10.1080/09639280902875556]
31. Topîrceanu A. Gamified learning: A role-playing approach to increase student in-class motivation. Procedia Comput Sci 2017;112:41-50. [doi: 10.1016/j.procs.2017.08.017]
32. Bastida L, Moya A, Gaeta E, de VF, José E, Gabler F. The power of gamification to learn and promote healthy habits among children. In: CEUR Workshop Proceedings. 2019 Presented at: CEUR Workshop; Nov 22, 2019; Barcelona, Spain.
33. Strmečki D, Bernik A, Radošević D. Gamification in e-learning: introducing gamified design elements into e-learning systems. J Comput Sci 2015 Dec 01;11(12):1108-1117 [FREE Full text] [doi: 10.3844/jcssp.2015.1108.1117]
34. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, CA: O'Reilly Media; Aug 22, 2011.
35. Ruipérez-Valiente J, Muñoz-Merino P, Delgado KC. Detecting and clustering students by their gamification behavior with badges: A case study in engineering education. Int J Eng Educ 2017;33(2-B):816-883 [FREE Full text]

XSL•FO
RenderX

36. Kelley HH, Michela JL. Attribution theory and research. Annu Rev Psychol 1980 Jan;31(1):457-501. [doi: 10.1146/annurev.ps.31.020180.002325] [Medline: 20809783]
37. Linehan C, Kirman B, Roche B. Gamification as behavioral psychology. In: Walz SP, Deterding S, editors. The Gameful World: Approaches, Issues, Applications. Cambridge, MA: MIT Press; 2015:81-105.

Abbreviations

EXP: experience points
MZ: millennials and Generation Z
PBL: point, badge, leaderboard

Edited by N Zary; submitted 22.12.21; peer-reviewed by O Ntsweng; comments to author 01.02.22; revised version received 08.03.22; accepted 11.08.22; published 29.09.22.

Please cite as:
Park S, Kim S
Points and the Delivery of Gameful Experiences in a Gamified Environment: Framework Development and Case Analysis
JMIR Serious Games 2022;10(3):e35907
URL: https://games.jmir.org/2022/3/e35907
doi:10.2196/35907
PMID:36173674

©Sungjin Park, Sangkyun Kim. Originally published in JMIR Serious Games (https://games.jmir.org), 29.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A New Approach for Reducing Virtual Reality Sickness in Real Time: Design and Validation Study

JuHye Won[1], MS; Yoon Sang Kim[2], PhD

[1]BioComputing Lab, Department of Computer Science and Engineering, Korea University of Technology and Education, Cheonan-si, Republic of Korea

[2]BioComputing Lab, Institute for Bio-engineering Application Technology, Department of Computer Science and Engineering, Korea University of Technology and Education, Cheonan-si, Republic of Korea

Corresponding Author:
Yoon Sang Kim, PhD
BioComputing Lab, Institute for Bio-engineering Application Technology
Department of Computer Science and Engineering
Korea University of Technology and Education
1600, Chungjeol-ro, Byeongcheon-myeon, Dongnam-gu
Cheonan-si, 31253
Republic of Korea
Phone: 82 0415601496
Email: yoonsang@koreatech.ac.kr

Abstract

Background: Recently, technology that provides virtual reality (VR) content based on streaming services has been rapidly developed. However, there have been few studies to reduce VR sickness that occurs while the user watches VR content while wearing a head-mounted display (HMD) in real time.

Objective: Based on this background, we propose a new approach to measure and reduce VR sickness that occurs while the user watches VR content while wearing an HMD in real time.

Methods: The proposed approach is to apply VR sickness reduction methods in accordance with the user's real-time VR sickness level. Three methods that are known to be effective in reducing VR sickness and a single type of VR content were used to examine the effectiveness of the proposed approach, which was confirmed by the experimental results.

Results: Our results show that VR sickness significantly decreased when a new approach was applied to VR content (in all cases, $P<.05$).

Conclusions: From our results, it was confirmed that VR sickness could be measured without wearing additional equipment, and its reduction method could be applied in real time in accordance with the user's condition by the proposed approach in this paper.

(JMIR Serious Games 2022;10(3):e36397) doi:10.2196/36397

KEYWORDS

virtual reality; VR; VR sickness; VR sickness reduction method; simulator sickness questionnaire; SSQ, visual guide; field of view; serious game; VR sickness reduction; VR content; technology; digital health

Introduction

Recently, technology that provides virtual reality (VR) content based on streaming services (such as YouTube VR, Netflix, etc) have been rapidly developed. Users of head-mounted displays (HMDs) are also increasing, and according to ResearchAndMarkets.com [1], the global HMD market is expected to grow by more than US $36 billion by 2026. However, VR sickness that occurs while the user watches VR content wearing an HMD does not have a positive effect on the proliferation of VR content. To solve this problem, research in the following 3 directions is being conducted.

The first direction is the identification of the cause of VR sickness. Several studies confirmed that speed [2], watching time [3], pitch, and roll rotation [4] affect VR sickness. Another study confirmed that the level of VR sickness varies depending on gender [5]. In addition, studies [6-8] were conducted to

XSL•FO
RenderX

confirm the correlation between VR sickness and VR content's elements, devices, and human factors.

The second direction is the derivation of the VR sickness measurement method. VR sickness measurement methods are divided into subjective and objective methods. Subjective methods were conducted through a survey such as simulator sickness questionnaires (SSQs) [9-11], the Motion Sickness Susceptibility Questionnaire (MSSQ) [12], Game Experience Questionnaire [13,14], Immersive Tendencies Questionnaire, and Presence Questionnaire [14]. Furthermore, in the objective method, studies were conducted to predict VR sickness using biological signals such as an electrocardiograms, electrodermal activity, electrooculogram, and breathing [9,10,15]. However, it is difficult to apply a VR sickness measurement method using biological signals to general users because additional equipment should be used. Furthermore, subjective measurement methods require surveys, and it is difficult to measure VR sickness in real time while VR content is being played.

The third direction is the derivation of VR sickness reduction methods. According to Singla et al [16] who conducted VR sickness reduction studies in terms of hardware, it was confirmed that HTC Vive provides an environment with a lower level of VR sickness than Oculus Rift. However, it is difficult to conclude that specific hardware is more effective in reducing VR sickness because new HMDs are constantly being developed. Therefore, studies are being conducted to reduce VR sickness in terms of content. A study was conducted to reduce VR sickness using the visual effects of VR content [17], and other studies have attempted to reduce VR sickness by applying a virtual human nose as an earth-fixed grid to the VR content [18,19]. A virtual human nose technique features the tip of a nose being fixed at the center-bottom of a VR user's view, acting as a rest-frame, which the brain can use to make natural spatial adjustments, thus reducing simulator sickness [18]. In addition, "Virtual Guiding Avatar" [20] combines various motion attributes with an independent visual background and Dynamic field of view (FOV) modification technology [21], which partially limits the user's FOV developed to reduce VR sickness. However, the above methods are only applied in advance in the content development step but do not apply VR sickness reduction methods in real time to suit the user's condition while VR content is being played.

As such, few studies have attempted to reduce VR sickness that occurs while the user watches VR content wearing an HMD in real time. Therefore, a new approach is proposed to measure and reduce VR sickness that occurs while the user watches VR content while wearing an HMD in real time. The proposed approach uses the VR sickness response, which is the result of the direct response of VR sickness that occurs while the user watches VR content wearing an HMD. Three methods that are known to be effective in reducing VR sickness have been used to examine the effectiveness of the proposed approach, which is confirmed by the experimental results. Furthermore, based on our results, the effect of the new approach for real-time VR sickness reduction and the effect of the VR sickness response are discussed.

Methods

A New Approach for VR Sickness Measurement and Reduction

This section describes a new approach for real-time VR sickness reduction. In the proposed approach, when VR sickness occurs while the user watches VR content wearing an HMD, the user clicks a button to directly express the VR sickness responses. In this paper, the response directly expressed by the user is defined as the "VR sickness response." Figure 1 shows a conceptual diagram of a new approach proposed to reduce VR sickness in real time. Users click a button when they experience VR sickness while wearing an HMD and watching VR content, and the sum of clicks becomes the VR sickness response. Then, the VR sickness reduction method is applied in accordance with the value of the VR sickness response. The following rules show how to apply a VR sickness reduction method in accordance with the VR sickness response:



For example, if the VR sickness response is equal to or greater than the threshold (α in the equation) within a certain the time interval (β in the equation), the VR sickness reduction method is applied. Furthermore, if the VR sickness response is less than the threshold (α in the equation) during a certain the time interval (β in the equation) while the VR sickness reduction method is applied, the VR sickness reduction method is not applied. The parameters required to apply the proposed approach, the time interval, and threshold are derived from a preliminary experiment as shown in Figure 2.

The subject clicks a button when he/she experiences VR sickness while watching VR content. In addition, whether the subject clicked the button every 1 second is saved (clicked, 1; nonclicked, 0). For example, if video playtime is a total of 60 seconds, 60 of the data points expressed as 0 or 1 are saved. Using this data set, the time interval and threshold are determined by the following 4 steps.

First, the sections of VR content are evenly divided in accordance with the time interval of several cases (eg, the time interval=5, 10, 15…n). Second, the average of the subject's VR sickness response in the evenly divided sections for each case is calculated. Third, the maximum and minimum values are selected from each case's average, and the difference is calculated. Finally, the time interval of the case with the smallest difference between the maximum and minimum values is set as the time interval of the proposed approach, and the same case's average VR sickness response is set as the threshold.

XSL•FO
RenderX

Figure 1. A conceptual diagram of a new approach proposed to reduce virtual reality (VR) sickness in real time. HMD: head-mounted display.

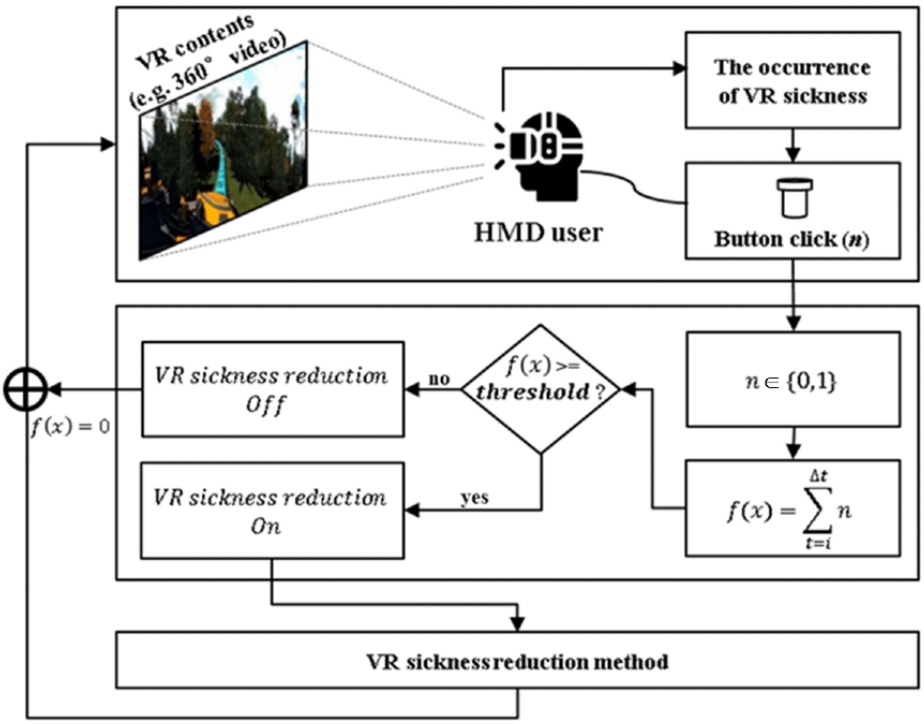
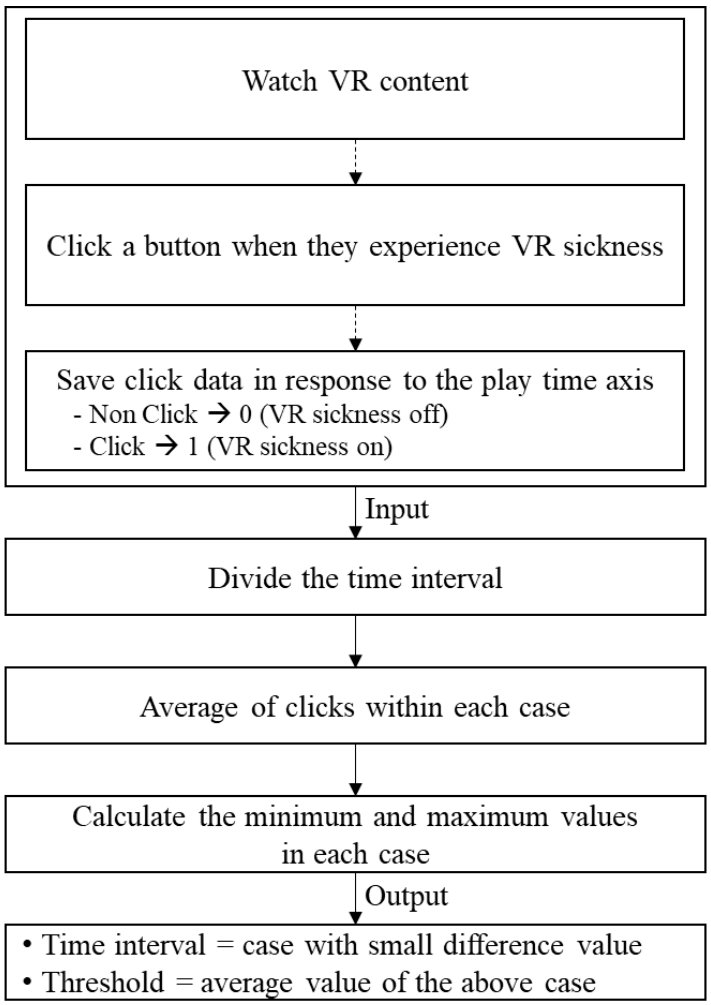


Figure 2. The Procedure for deriving the time interval and threshold. VR: virtual reality.



XSL•FO
RenderX

Experiments

This section describes the experimental method for confirming the effectiveness of the proposed new approach to reducing VR sickness in real time. The experiment was conducted on 40 subjects (20 male and 20 female) in their 20s and 30s. In addition, as shown in Figure 3, a 3D VR space flight video was selected as the original VR content and it was named M0.

A view of original VR content is changed in accordance with the subject's head movement (Figure 3B, which transitions to Figures 3A and 3C according to the head movement). Original VR content made subjects move the 3D space along a predetermined path for 60 seconds. In addition, acceleration, deceleration, and rotation (yaw, pitch, and roll) of the camera were applied to cause VR sickness. As shown in Figure 4, subjects watched VR content using HTC Vive Pro Eye, and they clicked the Xbox controller button to express their response to VR sickness in real time.

First, a preliminary experiment was conducted to derive the parameters of the proposed new approach (the time interval and threshold). In the preliminary experiment, VR sickness responses and SSQ for 40 subjects were measured. Figure 5 shows the time interval and threshold derived by the preliminary experiment. The original VR content used in the prior experiment were divided into a total of 4 cases in accordance with the time interval as shown in Figure 5A (the time interval of case 1 was 5 seconds; case 2, 10 seconds; case 3, 15 seconds; and case 3, 20 seconds). Then, in each case, the VR sickness response average value in the divided section was derived as shown in Figure 5B. As a result, the time interval of case 3, with the smallest difference between the maximum and minimum values of VR sickness response, was derived as the time interval of the proposed approach (time interval=15 seconds). In addition, the average value of VR sickness response in case 3 was derived as the threshold (threshold=3). After the parameters were derived, the VR sickness reduction method to be used in the proposed approach was selected. To this end, VR sickness reduction methods known to be effective were referred. Visual guide was a visual element that induces gaze movement and is known to be effective in reducing VR sickness [21]. In addition, it was known that the effect of VR sickness reduction is higher as the FOV decreases [22]. As a result, as shown in Table 1, a total of 3 VR sickness reduction methods were designed to be used in the proposed approach to reduce VR sickness in real time, and each method was named M1, M2, and M3.

The first VR sickness reduction method (M1) was applied with the visual guide, which had a 30% size of aspect ratio, and a position synchronized with the direction of the user's head movement. The second VR sickness reduction method (M2) applied the FOV that had a size of 90° and a position synchronized with the direction of the user's head movement.

When M2 is applied, if the subject moves their head from the center to the left of the screen, the FOV is limited on the basis of the left side of the screen.

The third VR sickness reduction method (M3) was applied, the FOV which had a size of 90° and a position synchronized to the direction of the user's gaze movement. When M3 is applied, if the subject moves their gaze from the center to the left of the screen, the FOV is limited on the basis of the left side of the screen.

Figure 6 shows examples of the VR sickness reduction methods to be used in the proposed new approach.

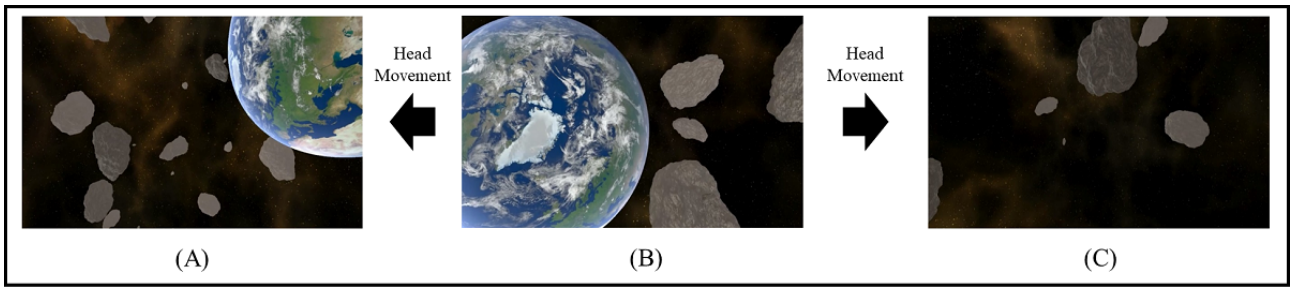
The experiment was performed using the protocol shown in Figure 7 for 3 VR sickness methods (M1-M3). The subject responded to MSSQ and SSQ before the experiment, and after wearing the HMD, calibration was performed for eye tracking. The experiment was conducted for a total of 35 min, and each of the methods (M1-M3) was randomly followed to ensure reliability. The subject's real-time VR sickness response was measured with respect to those 3 methods while watching VR content.

During the experiment, if the VR sickness response is measured more than 3 times within 15 seconds, the aforementioned 3 methods are applied.

Furthermore, the VR sickness reduction method is applied during the time interval, and if the VR sickness response is less than the threshold after the time interval, the method is discontinued.

After watching each VR content, subjects had time to respond to an SSQ and a questionnaire on fatigue and immersion; then, they had a period of rest to lower the level of VR sickness.

Figure 3. An example of the original virtual reality content: (A) and (C) are views changed in accordance with the user's head movement and (B) depicts the front view.



XSL•FO
RenderX

Figure 4. Experimental environment. HMD: head-mounted display; VR: virtual reality.

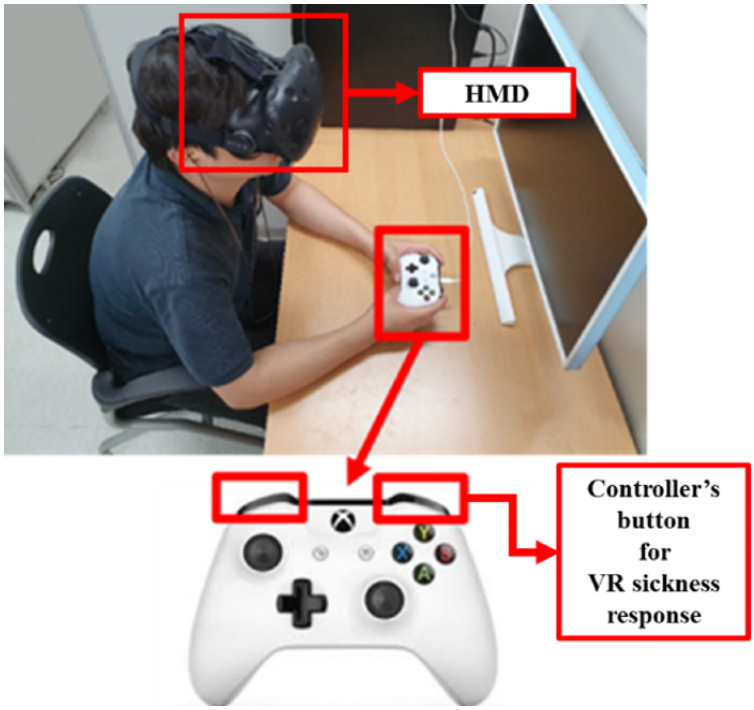
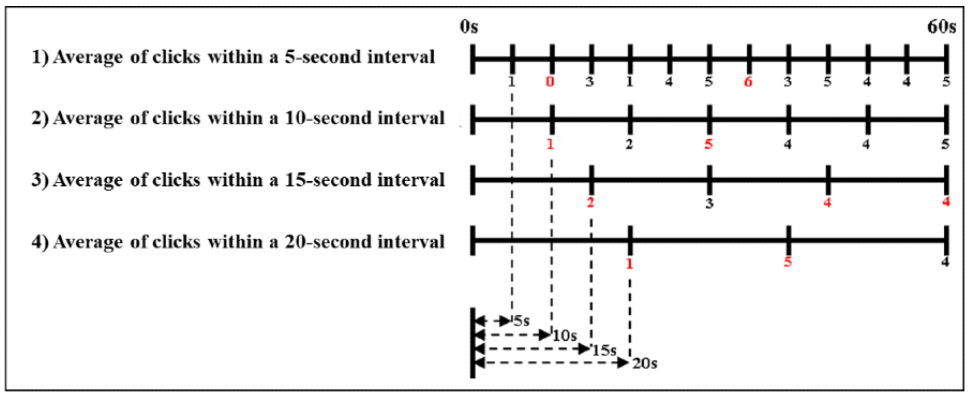


Figure 5. The time interval and threshold derived by the preliminary experiment: (A) are four cases of the preliminary experiment according to the time interval and (B) are the result of the preliminary experiment.

(A)



(B)

Time interval	Minimum value	Maximum value	Difference
5 seconds	0	6	6
10 seconds	1	5	4
15 seconds	2	4	2
20 seconds	1	5	4

- Time interval = 15 seconds
- Threshold = (2+3+4+4)/4 ≒ 3

Table 1. Three virtual reality sickness reduction methods were used for the proposed approach.

Method	Feature	Property
M1	Visual guide 1	• Size: 30% of aspect ratio • Position: movement with the direction of head movement
M2	Field of view 1	• Size: 90° • Position: movement with the direction of head movement
M3	Field of view 2	• Size: 90° • Position: movement with eye tracking (every second)

XSL•FO
RenderX

Figure 6. Examples of the virtual reality (VR) sickness reduction methods to be used in the proposed new approach (where, M0 is original VR content).

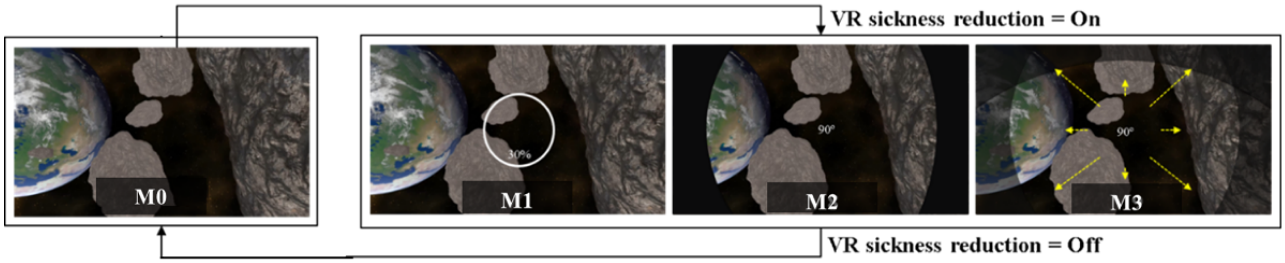
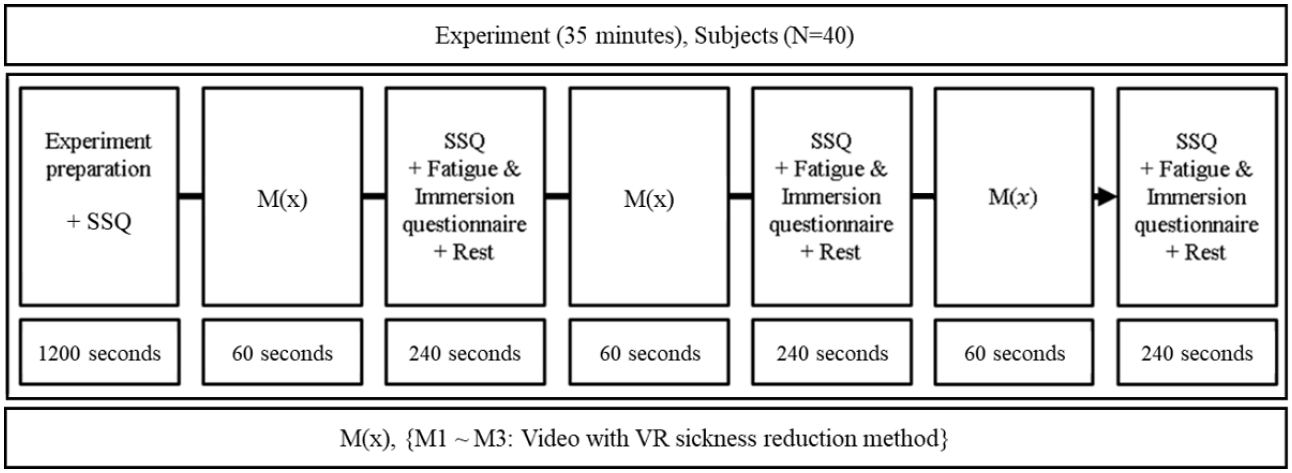


Figure 7. The experimental protocol. SSQ: simulator sickness questionnaire; VR: virtual reality.



Ethics Approval

This study was approved by the institutional review board of Korea University of Technology and Education (approval 19090401).

Results

This section describes the experimental results for confirming the effectiveness of the proposed new approach. In total, 26 out of 40 subjects did not experience VR sickness; hence, VR sickness reduction methods were not applied, and results based on 14 subjects who experienced VR sickness were analyzed. First, the SSQ scores of M0 and those of M1 to M3 were compared using a paired *t* test (2-tailed). As a result of the analysis, when VR sickness reduction methods (M1-M3) were used, the scores of nausea, oculomotor discomfort, and disorientation were significantly reduced (for all, $P<.05$).

The lowest SSQ score was observed in M2 and those of M1 and M3 were similar. Table 2 shows the results of VR sickness reduction in accordance with the SSQ score. Figure 8 shows the SSQ scores of the VR sickness reduction methods.

Furthermore, the VR sickness response of M0 and that of M1 to M3 were compared using a paired *t* test. All of the VR sickness reduction methods (M1-M3) showed significant VR sickness reduction.

Similar to the results of the SSQ scores, the lowest result was observed in M2, and the results of M1 and M3 were similar.

Table 3 shows the results of VR sickness reduction based on the VR sickness response. Figure 9 shows the VR sickness response to the methods.

Finally, from the results of the fatigue and immersion questionnaire, it was confirmed that M3 caused the most fatigue and simultaneously caused the lowest level of immersion. Table 4 shows the questionnaire results for fatigue and immersion. Figure 10 shows the questionnaire score for fatigue and immersion.

XSL•FO
RenderX

Table 2. The results of virtual reality sickness reduction method based on the simulator sickness questionnaire score.

Method	Nausea			Oculomotor discomfort			Disorientation			Total score		
	Score	*t* test (*df*)	*P* value	Score	*t* test (*df*)	*P* value	Score	*t* test (*df*)	*P* value	Score	*t* test (*df*)	*P* value
M0	47.02	—[a]	N/A[b]	44.40	—	N/A	79.54	—	N/A	61.71	—	N/A
M1	22.49	2.917[c] (13)	.01	21.66	3.261[d] (13)	.006	26.85	3.241[d] (13)	.006	26.71	3.230[d] (13)	.007
M2	15.67	3.175[b] (13)	.007	17.87	3.430[d] (13)	.004	22.87	3.277[d] (13)	.006	21.10	3.378[d] (13)	.005
M3	20.44	2.385[c] (13)	.03	23.28	2.668[c] (13)	.02	26.85	3.061[d] (13)	.009	26.71	2.766[c] (13)	.02

[a]Not determined.

[b]N/A: not applicable.

[c]Significant at $P<.05$.

[d]Significant at $P<.01$.

Figure 8. The simulator sickness questionnaire (SSQ) score of virtual reality (VR) sickness reduction methods (M).

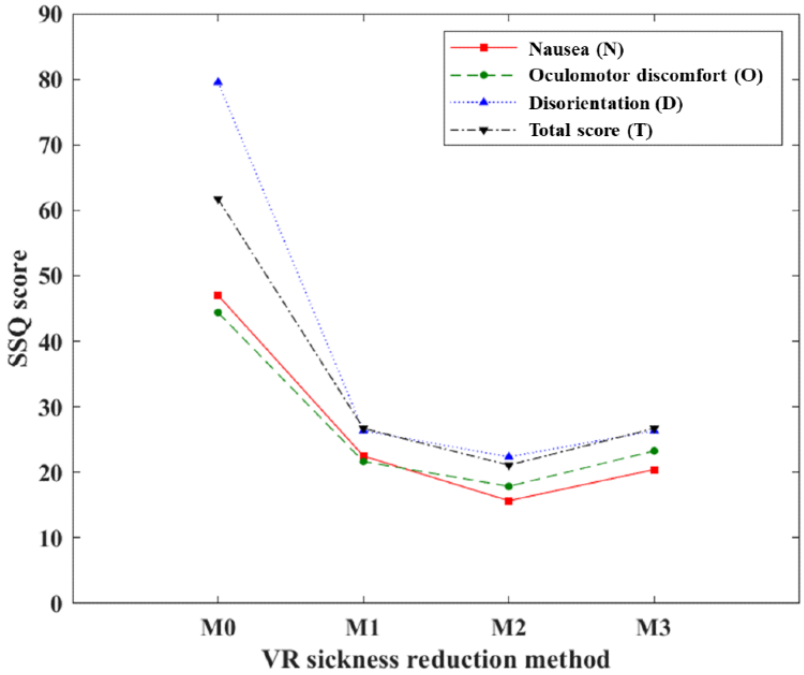


Table 3. The results of virtual reality (VR) sickness reduction based on the VR sickness response.

Real-time VR sickness measurement	Score	*t* test (*df*)	*P* value
M0	3.53	—[a]	N/A[b]
M1	1.66	4.988 (13)	<.001
M2	1.55	3.706 (13)	.003
M3	1.67	4.645 (13)	<.001

[a]Not determined.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 9. The average score of the virtual reality (VR) sickness response for the VR sickness reduction methods (M).

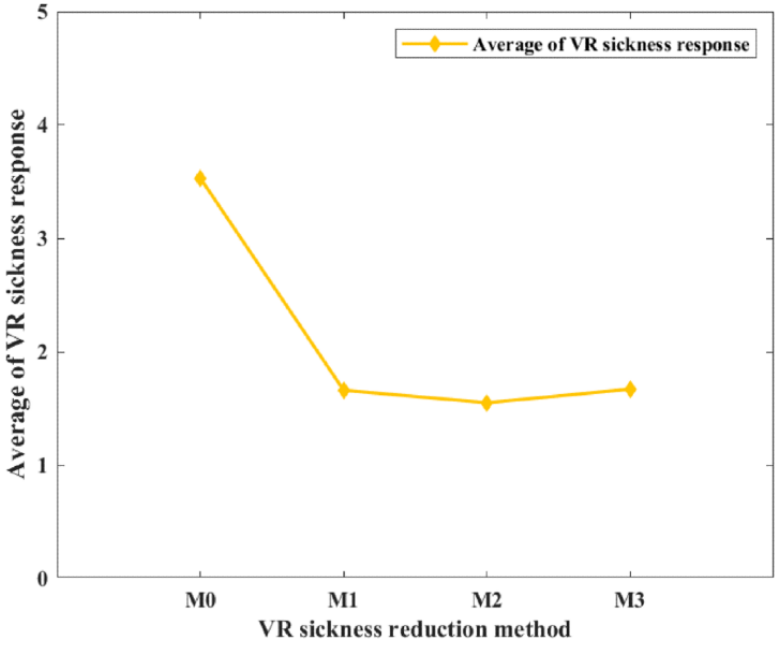
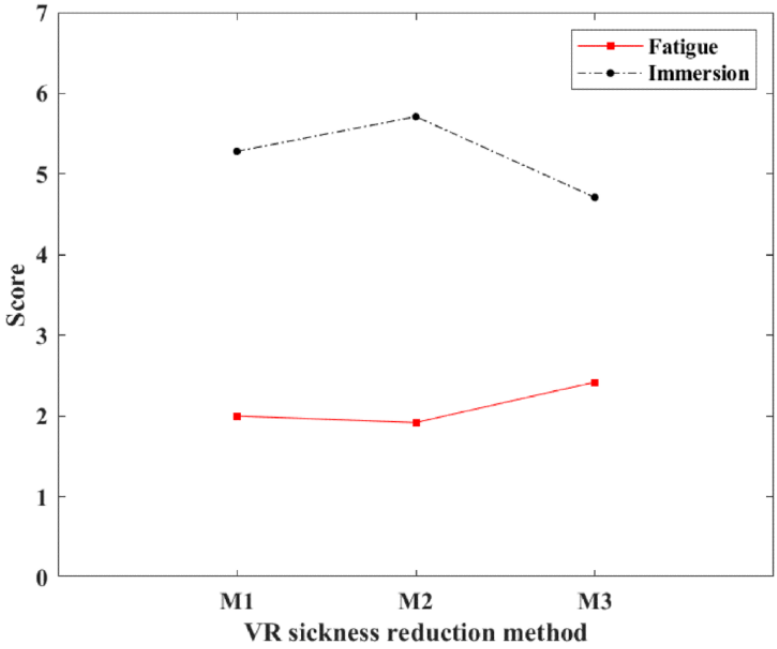


Table 4. The questionnaire results for fatigue and immersion.

Methods	Fatigue	Immersion
M1	2.00	5.28
M2	1.92	5.71
M3	2.42	4.71

Figure 10. The questionnaire score for fatigue and immersion. VR: virtual reality. M: VR sickness reduction methods.



Discussion

In this study, a new approach was proposed to reduce VR sickness in real time. The proposed approach used the VR sickness response, which is the result of the direct response of VR sickness that occurs while the user watches VR content wearing an HMD. Parameters necessary for the proposed approach were derived through preliminary experiments, and experiments were conducted to verify the effect of reducing VR sickness.

In addition, an experiment was conducted to confirm the effectiveness of the proposed approach, and from the experimental results, the following conclusions were drawn: (1) the methods (M1-M3) used previously [21,22] were effective

XSL•FO
RenderX

in reducing VR sickness; (2) the VR sickness response can indicate the user's VR sickness condition from the result that the subject's VR sickness response and SSQ score have a similar pattern; (3) a new approach to providing a VR sickness reduction method based on the VR sickness response was significant; and (4) the VR sickness reduction method synchronized with gaze movement, which caused fatigue and reduced immersion.

From our results, it was confirmed that VR sickness could be measured without wearing equipment, and the VR sickness reduction method could be applied in real time in accordance with the user's condition by the approach proposed in this paper. The proposed approach is expected to contribute to the spread of VR content by being applied to content streaming services.

This study has 2 limitations. First, a single type of VR content was used in the experiment; therefore, experiments using various types of VR content are needed to supplement the limitation. Second, the proposed method was applied after the user's VR sickness occurred; hence, further studies attempting to predict and reduce VR sickness are required.

Acknowledgments

This work was supported by the National Research Foundation of Korea grant funded by the Korea government (grant NRF-2020R1F1A1076114). This study was partially supported by the Education and Research Promotion Program of the Korea University of Technology and Education in 2021.

Conflicts of Interest

None declared.

References

1. Head-mounted Display Market with COVID-19 Impact Analysis by Type, Technology, Application, Product Type, Component, Connectivity and Geography - Global Forecast to 2026. GlobeNewswire. 2021 Jul 01. URL: https://tinyurl.com/3zjfu6dk [accessed 2022-09-15]
2. So RHY, Lo WT, Ho ATK. Effects of navigation speed on motion sickness caused by an immersive virtual environment. Hum Factors 2001;43(3):452-461. [doi: 10.1518/001872001775898223] [Medline: 11866200]
3. Porcino TM, Clua E, Trevisan D, Vasconcelos CN, Valente L. Minimizing cyber sickness in head mounted display systems: Design guidelines and applications. 2017 Presented at: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH); April 2-4, 2017; Perth, WA, Australia. [doi: 10.1109/segah.2017.7939283]
4. Joseph JA, Griffin MJ. Motion sickness: effect of the magnitude of roll and pitch oscillation. Aviat Space Environ Med 2008 Apr 01;79(4):390-396. [doi: 10.3357/asem.2196.2008] [Medline: 18457296]
5. Freitag S, Weyers B, Kuhlen TW. Examining rotation gain in CAVE-like virtual environments. IEEE Trans Visual Comput Graphics 2016 Apr 21;22(4):1462-1471. [doi: 10.1109/tvcg.2016.2518298]
6. Nie G, Yue L, Wang Y. Prevention of visually induced motion sickness based on dynamic real-time content-aware non-salient area blurring. 2017 Presented at: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct); October 9-13, 2017; Nantes. [doi: 10.1109/ismar-adjunct.2017.35]
7. Carnegie K, Rhee T. Reducing visual discomfort with HMDs using dynamic depth of field. IEEE Comput Grap Appl 2015 Sep;35(5):34-41. [doi: 10.1109/mcg.2015.98]
8. Benzeroual K, Allison RS. Cyber (motion) sickness in active stereoscopic 3D gaming. 2013 Presented at: 2013 International Conference on 3D Imaging; December 3-5, 2013; Liege. [doi: 10.1109/ic3d.2013.6732090]
9. Nesbitt K, Davis S, Blackmore K, Nalivaiko E. Correlating reaction time and nausea measures with traditional measures of cybersickness. Displays 2017 Jul;48:1-8. [doi: 10.1016/j.displa.2017.01.002]
10. Egan D, Brennan S, Barrett J, Qiao Y, Timmerer C, Murray N. An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. 2016 Presented at: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX); June 6-8, 2016; Lisbon. [doi: 10.1109/qomex.2016.7498964]
11. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
12. Kennedy R, Drexler JM, Compton DE, Stanney KM, Lanham DS, Harm DL. Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences?. 2003. URL: https://ntrs.nasa.gov/api/citations/20100033371/downloads/20100033371.pdf [accessed 2022-09-16]
13. Stanney KM, Hale KS, Nahmens I, Kennedy RS. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. Hum Factors 2003 Aug 28;45(3):504-520. [doi: 10.1518/hfes.45.3.504.27254] [Medline: 14702999]
14. Haans A, IJsselsteijn WA, de Kort YAW. Self-Report Measure [Database record]. APA PsycTests. 2008. URL: https://doi.org/10.1037/t21169-000 [accessed 2022-09-16]

XSL•FO
RenderX

15. Dennison MS, Wisti AZ, D'Zmura M. Use of physiological signals to predict cybersickness. Displays 2016 Sep;44:42-52. [doi: 10.1016/j.displa.2016.07.002]
16. Singla A, Fremerey S, Robitza W, Raake A. Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. 2017 Presented at: 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX); May 31 to June 2, 2017; Erfurt. [doi: 10.1109/qomex.2017.7965658]
17. Buhler H, Misztal S, Schild J. Reducing VR sickness through peripheral visual effects. 2018 Presented at: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); March 18-22, 2018; Tuebingen/Reutlingen. [doi: 10.1109/vr.2018.8446346]
18. Wienrich C, Weidner CK, Schatto C, Obremski D, Israel JH. A virtual nose as a rest-frame - the impact on simulator sickness and game experience. 2018 Presented at: 2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games); September 5-7, 2018; Würzburg. [doi: 10.1109/vs-games.2018.8493408]
19. Whittinghill DM, Ziegler B, Case T, Moore B. Nasum Virtualis: A Simple Technique for Reducing Simulator Sickness in Head Mounted VR. 2015 Mar 2 Presented at: Games developers conference (GDC); March 2-6, 2015; San Francisco p. A.
20. Lin JJW, Abi-Rached H, Lahav M. Virtual guiding avatar: an effective procedure to reduce simulator sickness in virtual environments. 2004 Presented at: CHI04: CHI 2004 Conference on Human Factors in Computing Systems; April 24-29, 2004; Vienna. [doi: 10.1145/985692.985783]
21. Won J, Kim YS. A study on visually induced VR reduction method for virtual reality sickness. Applied Sciences 2021 Jul 08;11(14):6339. [doi: 10.3390/app11146339]
22. Fernandes AS, Feiner SK. Combating VR sickness through subtle dynamic field-of-view modification. 2016 Presented at: 2016 IEEE Symposium on 3D User Interfaces (3DUI); March 19-20, 2016; Greenville, SC. [doi: 10.1109/3dui.2016.7460053]

Abbreviations

FOV: field of view
HMD: head-mounted display
MSSQ: Motion Sickness Susceptibility Questionnaire
SSQ: simulator sickness questionnaire
VR: virtual reality

Edited by N Zary; submitted 13.01.22; peer-reviewed by B Concannon, J Lebiedz'; comments to author 28.04.22; revised version received 09.06.22; accepted 04.08.22; published 27.09.22.

Please cite as:
Won J, Kim YS
A New Approach for Reducing Virtual Reality Sickness in Real Time: Design and Validation Study
JMIR Serious Games 2022;10(3):e36397
URL: https://games.jmir.org/2022/3/e36397
doi:10.2196/36397
PMID:36166294

©JuHye Won, Yoon Sang Kim. Originally published in JMIR Serious Games (https://games.jmir.org), 27.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Effect of the "Art Coloring" Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation

JuZhe Xi[1*], PhD; YuHan Gao[1*], MSc; Na Lyu[2], MSc; Zhuang She[1], MSc; XinYue Wang[3], BA; Xin-An Zhang[4], PhD; XiaoYu Yu[5], PhD; WeiDong Ji[6], MD; MengSheng Wei[7], MA; WeiHui Dai[7], PhD; XueSheng Qian[7], PhD

[1]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

[2]Xinhua Hospital, Shanghai Jiao Tong University of Medicine, Shanghai, China

[3]College of Letters and Science, University of California, Berkeley, CA, United States

[4]Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai, China

[5]School of Management, Shanghai University, Shanghai, China

[6]Affiliated Mental Health Center (ECNU), Shanghai Changning Mental Health Center, Shanghai, China

[7]School of Management, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
XueSheng Qian, PhD
School of Management
Fudan University
Room 722, Siyuan Building
670 Guoshun Road
Shanghai, 200433
China
Phone: 86 13916208229
Email: qianxuesheng@fudan.edu.cn

Related Article:

Correction of: http://mhealth.jmir.org/2022/3/e37026/

(JMIR Serious Games 2022;10(3):e41253) doi:10.2196/41253

In "Effect of the "Art Coloring" Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation" (JMIR Serious Games 2022;10(3):e37026) the authors made changes to affiliations 1 and 6:

1. In the originally published article, affiliation 1 was the following:

> *Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China*

A subunit was added to affiliation 1, which now reads as follows:

> *Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China*

2. In the originally published article, affiliation 6 was the following:

> *Shanghai Changning Mental Health Center, East China Normal University, Shanghai, China*

A change was made to affiliation 6, which now reads as follows:

> *Affiliated Mental Health Center (ECNU), Shanghai Changning Mental Health Center, Shanghai, China*

The correction will appear in the online version of the paper on the JMIR Publications website on August 16, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 21.07.22; this is a non–peer-reviewed article;accepted 25.07.22; published 16.08.22.

Please cite as:
Xi J, Gao Y, Lyu N, She Z, Wang X, Zhang XA, Yu X, Ji W, Wei M, Dai W, Qian X
Correction: Effect of the "Art Coloring" Online Coloring Game on Subjective Well-Being Increase and Anxiety Reduction During the COVID-19 Pandemic: Development and Evaluation
JMIR Serious Games 2022;10(3):e41253
URL: https://games.jmir.org/2022/3/e41253
doi:10.2196/41253
PMID:35973213

©JuZhe Xi, YuHan Gao, Na Lyu, Zhuang She, XinYue Wang, Xin-An Zhang, XiaoYu Yu, WeiDong Ji, MengSheng Wei, WeiHui Dai, XueSheng Qian. Originally published in JMIR Serious Games (https://games.jmir.org), 16.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Using Augmented Reality for Training in the Medical Professions: Meta-analysis

Yahia Baashar[1], MSc; Gamal Alkawsi[2], PhD; Wan Nooraishya Wan Ahmad[1], PhD; Hitham Alhussian[3], PhD; Ayed Alwadain[4], PhD; Luiz Fernando Capretz[5], PhD; Areej Babiker[6], PhD; Adnan Alghail[7], PhD

[1]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Labuan, Malaysia

[2]Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

[3]Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Malaysia

[4]Department of Computer Science, King Saud University, Riyadh, Saudi Arabia

[5]Department of Electrical & Computer Engineering, Western University, Ontario, ON, Canada

[6]Department of Computer Engineering, Future University, Khartoum, Sudan

[7]Department of World Languages, Greece Central School District, New York, NY, United States

Corresponding Author:
Yahia Baashar, MSc
Faculty of Computing and Informatics
Universiti Malaysia Sabah
751 Bandar
Labuan, 88400
Malaysia
Phone: 60 0102207514
Email: yahia.baashar@gmail.com

Abstract

Background: Augmented reality (AR) is an interactive technology that uses persuasive digital data and real-world surroundings to expand the user's reality, wherein objects are produced by various computer applications. It constitutes a novel advancement in medical care, education, and training.

Objective: The aim of this work was to assess how effective AR is in training medical students when compared to other educational methods in terms of skills, knowledge, confidence, performance time, and satisfaction.

Methods: We performed a meta-analysis on the effectiveness of AR in medical training that was constructed by using the Cochrane methodology. A web-based literature search was performed by using the Cochrane Library, Web of Science, PubMed, and Embase databases to find studies that recorded the effect of AR in medical training up to April 2021. The quality of the selected studies was assessed by following the Cochrane criteria for risk of bias evaluations.

Results: In total, 13 studies with a total of 654 participants were included in the meta-analysis. The findings showed that using AR in training can improve participants' performance time (I^2=99.9%; P<.001), confidence (I^2=97.7%; P=.02), and satisfaction (I^2=99.8%; P=.006) more than what occurs under control conditions. Further, AR did not have any effect on the participants' knowledge (I^2=99.4%; P=.90) and skills (I^2=97.5%; P=.10). The meta-regression plot shows that there has been an increase in the number of articles discussing AR over the years and that there is no publication bias in the studies used for the meta-analysis.

Conclusions: The findings of this work suggest that AR can effectively improve performance time, satisfaction, and confidence in medical training but is not very effective in areas such as knowledge and skill. Therefore, more AR technologies should be implemented in the field of medical training and education. However, to confirm these findings, more meticulous research with more participants is needed.

(JMIR Serious Games 2022;10(3):e32715) doi:10.2196/32715

KEYWORDS

augmented reality; medical; training; virtual; meta-analysis

XSL•FO
RenderX

Introduction

Augmented reality (AR) is an interactive technology that uses persuasive digital data and real-world surroundings to expand the user's reality, wherein objects are produced by various computer applications [1,2]. Its use has become a widely debated topic among the medical community and has resulted in fertile ground for new types of research [3]. As a result, the health care sector and policy makers have been immediately able to realize the benefits of applying AR technologies, with education and training being noticeable applications of AR in medical settings [4-6]. However, since AR technologies enable learners to interact and visualize in 3D representations, medical professionals must have a vast amount of technological knowledge to efficiently adapt AR in practice [4,7,8]. Apart from the enormous benefits of AR for medical professionals, AR is also emerging as a useful guide for patient education and training, allowing physicians to clarify different surgical procedures and the functions of certain medications for patients and to explain new therapies in a more visual way [9-12]. Furthermore, AR can significantly improve a patient's experience and advance knowledge on complex issues, and these benefits contribute and lead to better health outcomes [6].

There are many applications of AR in medical settings. Kamphuis et al [13], for example, reported that AR has reached a mature level of use in the field of anatomical engineering, for various technologies, and in physiological training. Further, surgical training has also been revolutionized by AR. Although integrating virtual reality (VR) and AR training into medical school and residency programs would necessitate prototype testing and entirely new teaching methods, AR has the potential to completely transform medical and science training [14-16]. Early adopters of AR innovations in medical training have noted several benefits, including the ability to simulate real-life scenarios without risking real-world consequences. Despite these advancements, there is still a lack of published experimental and observational work on the effectiveness and usefulness of AR in medical training and education [16].

In their work, Eckert et al [17] provided an overview of the development of AR applications in health care from 2012 to 2017. According to their findings, despite an increasing number of publications on AR in medicine, there have been no significant clinical trials on AR effectiveness. However, display-related domains tended to receive more research attention than other AR interventions.

The literature shows some discrepancies in the effectiveness of AR in medical training. To date, most of the meta-analysis research on the efficiency of VR and virtual patients in medical education has been carried out by Chen et al [18] and Kononowicz et al [19]. Their research has inspired this study, and there is some thematic overlap, but they presented their findings in a different way. However, to the best of our knowledge, no meta-analysis has been conducted to assess AR in medical training, with the exception of 2 studies—those by Williams et al [20] and Barsom et al [16], which conducted a systematic review with no meta-analysis. Therefore, filling this gap represents a significant addition to the literature. The primary goal of this study is to perform a meta-analysis on the efficacy of AR in 5 areas of learning and training. These include participants' skills (eg, the ability to illustrate a certain process), knowledge (eg, the ability to understand a certain concept), confidence (eg, the ability to learn certain content with self-confidence), performance time (eg, the amount of time spent on a certain task), and satisfaction (eg, the level of learning fulfilment for a certain AR intervention). The findings of this work have many implications for policy makers, patients, medical students, and professionals and for using AR technology to facilitate learning and training mechanisms in the medical field. This meta-analysis seeks to address the following research questions (RQs):

- RQ1: What are the characteristics of studies using AR in medical training?
- RQ2: What kind of AR interventions were used to assist with medical training?
- RQ3: Does the use AR have an effect on medical training when compared to other methods?

Methods

Overview

The *Methods* section discusses the approaches used to obtain specific publications as well as the constraints and eligibility requirements that were used [21]. Following that, an overview of the standards used to interpret the studies and associated variables is provided. This meta-analysis was carried out in compliance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [22]. A checklist of all PRISMA items is presented in Multimedia Appendix 1.

Eligibility Criteria

The meta-analysis included studies that (1) were published between 2013 and April 2021, (2) were written in English, and (3) primarily focused on AR for learning and training. Further, we excluded studies that (1) used other virtual platforms, such as VR and virtual patients, and (2) had no experimental data. However, based on the PICO (Population, Intervention, Comparison, Outcomes) framework, our final analysis incorporated various trials, including those with randomized controlled and mixed designs. Textbox 1 illustrates all of the PICO elements used for the inclusion criteria.

XSL•FO
RenderX

Textbox 1. PICO (Population, Intervention, Comparison, Outcomes) framework.

Criteria and descriptions

- Population
  - Sex (male or female)
  - Age (>15 years)
  - Target group (medical professionals, residents, patients, or students)
- Intervention
  - All types of augmented reality technology used in training or learning.
- Comparison
  - Augmented reality, nonaugmented reality, and conventional methods (eg, paper or pen, classes, simulations, or presentations)
- Outcomes
  - Knowledge
  - Confidence
  - Skills
  - Performance time
  - Satisfaction

Information Source and Search Strategy

We conducted a web-based literature search in the Cochrane Library, PubMed, Embase, Web of Science and Google Scholar databases from February to April 2021. We also manually checked the reference lists of the eligible studies to identify and acquire additional relevant publications. We used the following set of keywords, terms, and logical operators: "augmented reality" OR "AR" OR "mixed reality" OR "patient simulation" AND "medical training" OR "medical education" OR "health professions training." Some of the search strategies and terms used are presented in Multimedia Appendix 2. It is worth mentioning that the search strategies were slightly modified to suit each web-based database. EndNote X9 (Clarivate) software was used to manage and import the selected documents and to remove duplicates. This process was carried out by 2 authors (YB and GA) of this study.

Data Extraction

Data were placed into an extraction spreadsheet by using Microsoft Excel 2019. From each of the studies, information was extracted based on the study characteristics, designs, and participants. These data included the names of authors, the years of the studies, the locations of the studies, sample sizes, participant types, study designs, and interventions and outcomes. EndNote X9 was also used to obtain some publication data, such as titles, publishers, URLs, digital object identifiers, page numbers, issue numbers, and volume numbers. The process of data extraction was initially performed by 2 authors (HA and A Alwadain) and was then validated by another 2 authors (WNWA and LFC).

Risk of Bias Assessment

The quality of the studies was assessed by 2 authors (AB and A Alghail). We followed the criteria of the Cochrane Handbook for Systematic Reviews of Interventions [23] in the risk of bias assessment process. These include 7 domains that correspond to a particular kind of bias. Each domain was given a score (a "high risk," "low risk," or "unclear risk" of bias), and any disagreements among coauthors were settled via consensus.

Statistical Analysis

RevMan 5.4 (The Cochrane Collaboration) software was used to perform the risk of bias assessment; different scales were used by different authors to measure the outcomes; therefore, raw data could not be compared directly. Thus, the R 4.0.2 (R Foundation for Statistical Computing) software was used to transform raw data, so that STATA 14.0 (StataCorp LLC) could be used to run the analysis directly. The *metacont* package on R was used to transform the data on the sample sizes, SDs, and means of the control and AR groups. Data on the sample sizes, means, and SDs of only the AR groups were transformed by using the *metamean* package. Only 1 study [24] recorded the sample size and the proportion of participants on whom AR had a positive effect. The effect sizes and CIs were generated by using the two different packages. STATA 14.0 was used to generate the forest plot for each outcome, which used the transformed data (effect size and CI), via the *metan* command. The statistical heterogeneity among the selected studies was measured by using I^2 in each analysis and a 5% significance level. The random effect model was selected for the analysis carried out because the true effect sizes underlying all studies were stochastic.

Results

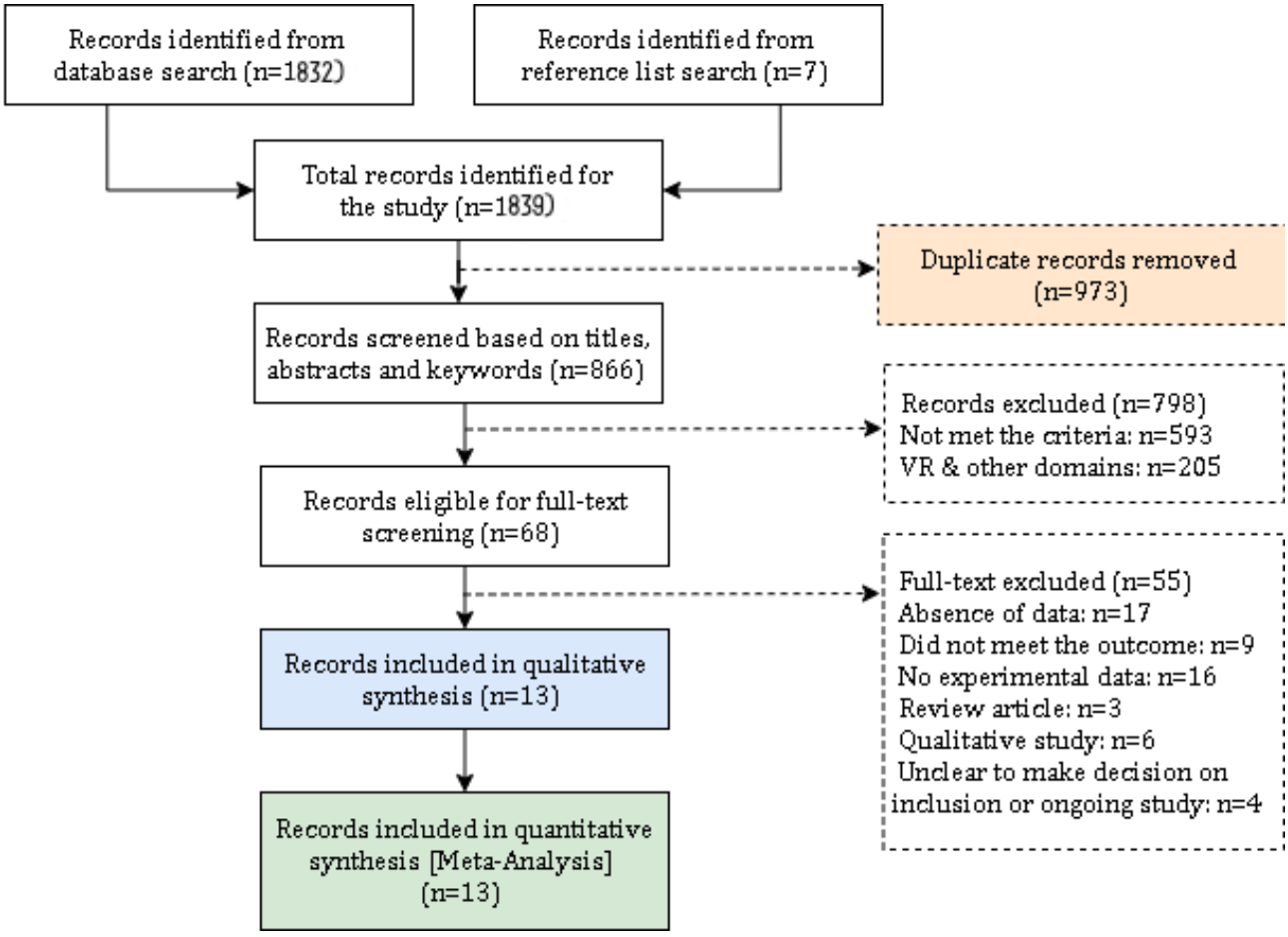
Study Selection

From our initial search of 5 potential databases, we found a total of 1832 records. We identified a total of 1839 records, including

XSL•FO
RenderX

relevant reference lists (n=7). Duplicate records (973/1839, 52.9%) were removed. Following that, a total of 866 of the 1839 (47.1%) studies were screened based on their titles, abstracts, and keywords, and 798 were excluded after applying the eligibility criteria. A total of 68 studies were downloaded and evaluated via full-text screening, of which 55 were excluded for various reasons (Figure 1). Finally, 13 studies with a total of 654 participants were eligible for the meta-analysis; 2 studies used the same set of participants for both the control and AR groups [25,26], while another 6 studies [27-32] that also measured both the control and AR groups divided the sample size.

Figure 1. Study screening and selection flowchart. VR: virtual reality.



Study Characteristics

As presented in Table 1, the trials of the selected studies were performed in the following ten countries: Germany [27,28,30], the United States [24,31], the United Kingdom [33], Canada [34], Italy [35], Sweden [26], Finland [32], Switzerland [29], Spain [36], and South Korea [25]. Furthermore, 10 of the 13 studies, adopted a randomized control trial [24,26-29,32-36], 2 studies had a mixed design [30,31], and 1 study used a cohort approach [25]. The number of participants in the studies ranged between 4 and 372. With regard to these participants, 5 trials recruited medical and nursing students [25,27,28,30,31]; 2 trials involved pediatric and psychiatry residents [26,29]; 1 trial recruited visitors to a center [32]; and the remaining trials (n=5) included multiple participants, such as physicians, registered nurses, technicians, residents, students, clinicians, clients, hosts, and paramedics [24,33-36].

In recent years, the use of AR in the medical field has attracted the attention of academics and researchers, as evidenced by the increasing number of publications devoted to AR. This is demonstrated by the fact that 8 of the 13 studies were conducted between 2019 and 2021. The characteristics of the selected studies, including the outcome measures, interventions, and types of participants, are shown in Table 1. Additional details regarding the AR and control groups are presented in Multimedia Appendix 3.

XSL•FO
RenderX

Table 1. Characteristics of the selected studies.

Author, year	Country	Brief description	Participants	Sample size, n (number of participants in each study group)	Study design	Outcomes
Albrecht et al, 2013 [27]	Germany	Comparing the effect of MAR[a] to that of a textbook, with consideration for essential psychological qualities	Students	10 (6 and 4)	RCT[b]	Skill and confidence
Balian et al, 2019 [24]	United States of America	Testing the feasibility of an AR[c] training system (Microsoft HoloLens) for CPR[d] among medical professionals	Physicians, nurses, and technicians	51 (N/A[e])	RCT	Performance time and satisfaction
Ingrassia et al, 2020 [35]	Italy	Assessing the feasibility of an AR prototype for BLSD[f] training	Physicians, nurses, and residents	26 (N/A)	RCT	Confidence and satisfaction
Kim et al, 2021 [25]	Korea	Evaluating the usability and feasibility of AR (smart glasses) for nursing training skills	Students	30 (N/A)	Cohort	Skill, performance time, satisfaction, and knowledge
Kotcherlakota et al, 2020 [31]	United States of America	Assessing the use of AR (clinical simulation) in the management of pediatric asthma outcomes	Students	21 (12 and 9)	Mixed	Confidence and satisfaction
Muangpoon et al, 2020 [33]	United Kingdom	Proposing an AR system (Microsoft *HoloLens, 1st Gen*) to improve the learning and teaching of DRE[g]	Clinicians and students	19 (N/A)	RCT	Skill, knowledge, and satisfaction
Noll et al, 2017 [28]	Germany	Assessing learning success by comparing learners with and without MAR	Students	44 (22 and 22)	RCT	Knowledge and skill
Pantziaras et al, 2015 [26]	Sweden	Evaluating the impact of virtual patient training on the knowledge of stress disorder management and symptoms	Residents	32 (N/A)	RCT	Knowledge
Savela et al, 2020 [32]	Finland	Investigating the features of MAR for learning and sociability	Visitors	372 (231, 71, and 71)	RCT	Knowledge and satisfaction
Schiffeler et al, 2019 [30]	Germany	Assessing the effects of AR on interaction and communication	Students	13 (7 and 6)	Mixed	Knowledge
Siebert et al, 2017 [29]	Switzerland	Evaluating whether the adaption of AR glasses with AHA[h] guidelines can reduce the time and deviation of essential life-saving exercises throughout pediatric CPR when compared to those of PALS[i]	Residents	20 (10 and 10)	RCT	Performance time and confidence
Vidal-Balea et al, 2021 [36]	Spain	Evaluating MAR games that teach and train people how to use pediatric medical applications (the games also monitor training progress)	Clients and hosts	4 (N/A)	RCT	Performance time
Wang et al, 2017 [34]	Canada	Developing a telemedicine platform using AR (Microsoft HoloLens) to improve medical training remotely	Paramedics and students	12 (N/A)	RCT	Performance time and satisfaction

[a]MAR: mobile augmented reality.

[b]RCT: randomized controlled trial.

[c]AR: augmented reality.

[d]CPR: cardiopulmonary resuscitation.

[e]N/A: not applicable.

[f]BLSD: basic life support and defibrillation.

[g]DRE: digital rectal examination.

[h]AHA: American Heart Association.

[i]PALS: pediatric advanced life support.

XSL•FO
RenderX

Risk of Bias

Figure 2 and Figure 3 present the risk of bias assessment, which was performed according to the Cochrane criteria. Of the 13 studies, 10 detailed randomized methods [24,26-28,30-33,35,36], while the rest of the studies (n=3) [25,29,34] reported no sequence generation methods. Only 3 trials reported concealment methods [25,32,34]. Furthermore, 3 studies reported dropouts [29,30,33] but did not go into detail about how they were handled. Due to the uniqueness of the intervention methods, no blinding methods were reported for participants in any of the trials. The blinding of assessors was used in 5 trials [26,27,29,35,36].

Figure 2. Risk of bias assessment of each selected study [24-36].

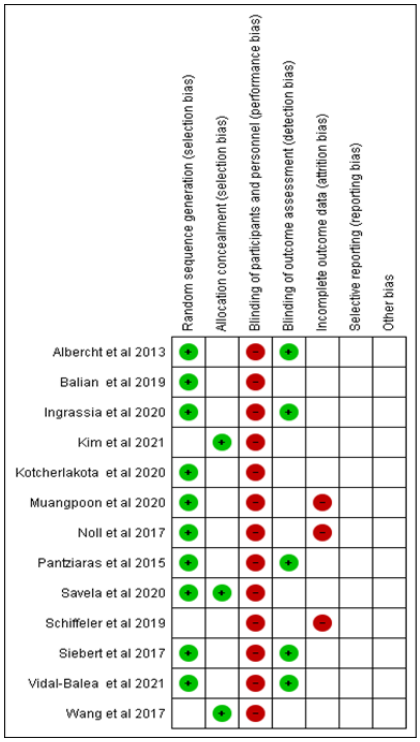
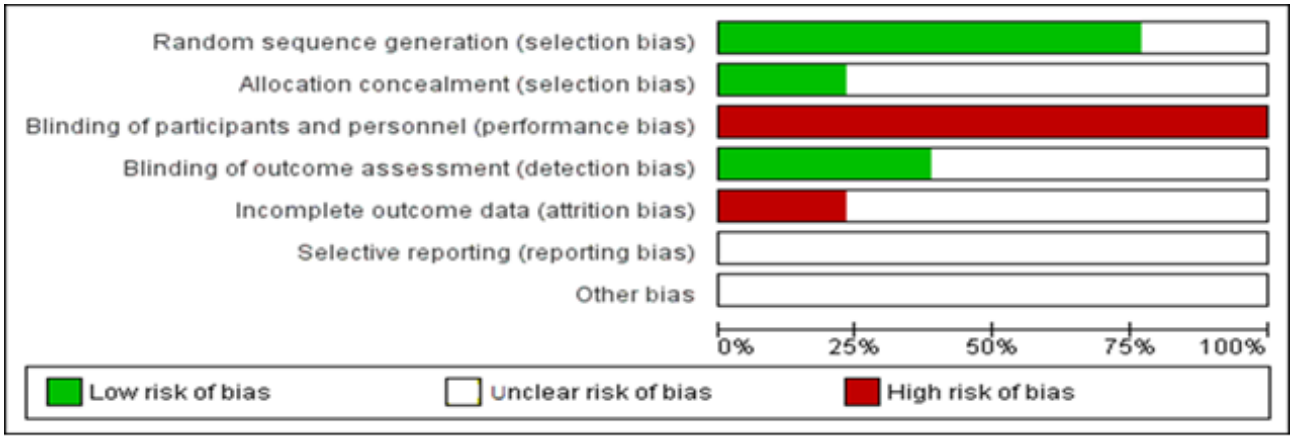


Figure 3. Overall risk of bias assessment of the selected studies.



Publication Bias

The most significant issue with meta-analyses is the possibility of bias and error arising from combining multiple studies, especially unrelated studies. Creating a funnel plot is one of the key methods used for testing common publication bias. The graphics in the funnel are obvious scatter plots for the effect size approximated in each selected study versus the unit of sample size reported in the studies [37]. If there is no bias, the graphic resembles an upside-down funnel. However, if a publication bias exists, the graphic should be asymmetric and distorted [38]. Figure 4 shows the funnel plot that we used to test for publication bias.

As can be seen in Figure 4, the 13 studies included in the meta-analysis are symmetrically placed on both sides of the vertical effect size line and are very close to the symmetrical and merged effect size lines. If there is no publication bias, studies should be distributed symmetrically on both sides of the vertical line depicting the merged effect size. If the studies are placed outside of the pyramid, they should be concentrated in the upper or middle parts of the plot. However, if there is a bias, most of the studies will be in the funnel plot's bottom part or in

XSL•FO
RenderX

only 1 vertical line segment [39]. The effect sizes of the included studies are distributed in a symmetrical manner, with only 5 studies located outside of the pyramid. However, it is worth nothing that 4 of these studies are located in the upper part of the pyramid. This pattern shows that the studies included in this work have no publication bias.

As shown in Figure 5—the meta-regression plot—the true effect sizes between the control and AR groups are 1.8 for the 2013 study [27]; 0.9 for the 2015 study [26]; 2.2, 0.2, and −0.9 for the 2017 studies [28,29,34], respectively; 1.2 and 0.9 for the 2019 studies [24,30], respectively; 0.4, −1, 0.2, and 1.5 for the 2020 studies [31-33,35], respectively; and finally, −3.1 and 0.8 for the 2021 studies [25,36], respectively.

Figure 4. Funnel plot showing publication bias.

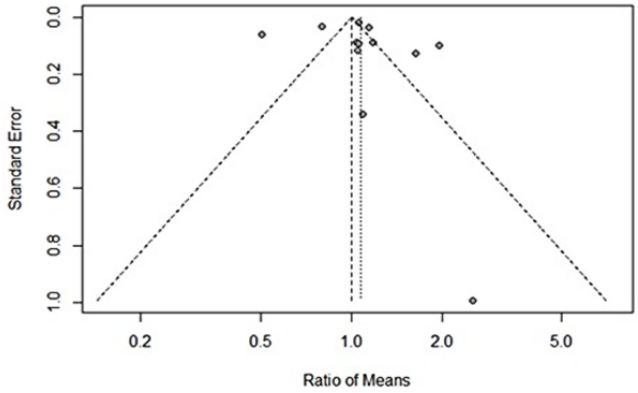
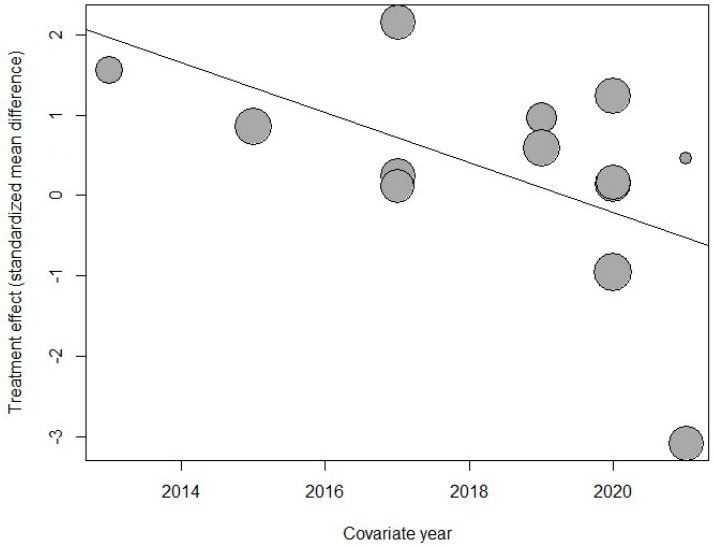


Figure 5. Meta-regression plot showing the publication years.



Results of the Meta-analysis

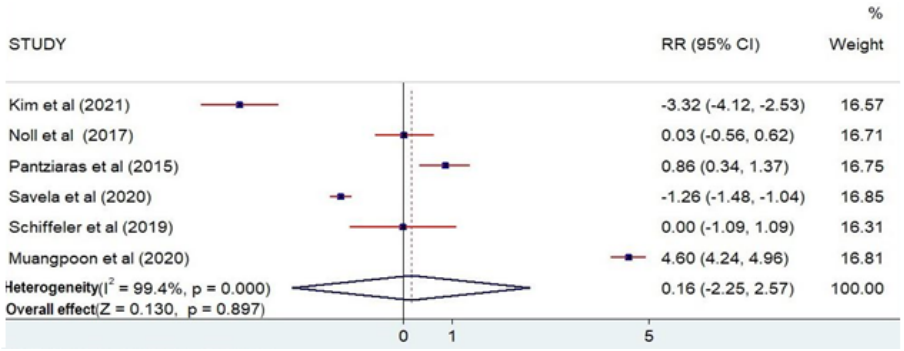
A meta-analysis was performed by using the data set in Multimedia Appendix 4 for each of the five outcomes (ie, knowledge [Multimedia Appendix 5], performance time [Multimedia Appendix 6], satisfaction [Multimedia Appendix 7], confidence [Multimedia Appendix 8], and skills [Multimedia Appendix 9]). The coding of the final analysis is also provided in Multimedia Appendix 10.

Knowledge

Only 6 of the 13 selected studies recorded knowledge as scores [25,26,28,30,32,33]. The total number of participants who measured their knowledge was 510. The findings from the plot (Figure 6) showed that there is no relationship between AR training and improvement in the participants' knowledge (Z=0.130; P=.90).

XSL•FO
RenderX

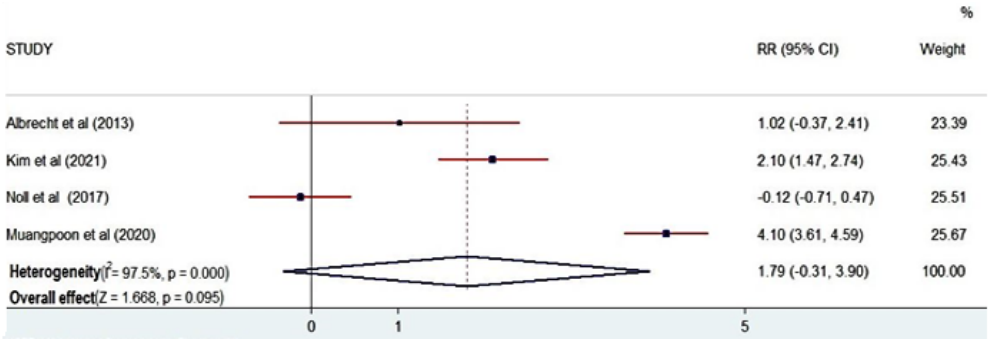
Figure 6. Forest plot showing the effectiveness of augmented reality on knowledge [25,26,28,30,32,33]. Weights are from the random-effects model. RR: risk ratio.



Skills

A total of 4 studies used skill as the outcome measurement [25,27,28,33]. The total number of participants that participated in measuring skill was 103. The plot (Figure 7) revealed that there is no statistical relationship between AR training and participants' skills (Z=1.668; P=.10). There is high heterogeneity among the studies.

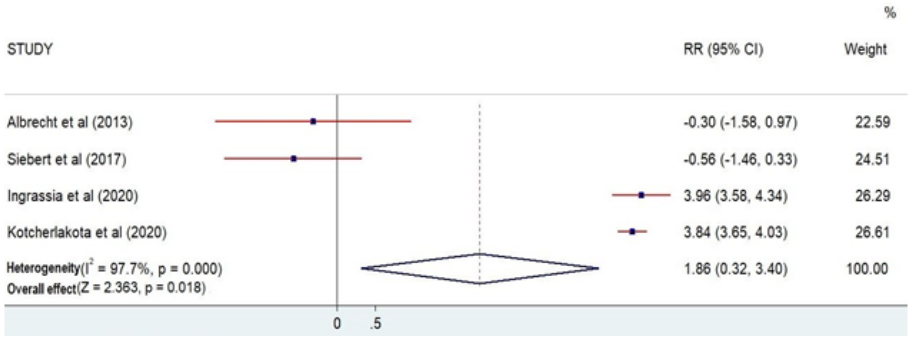
Figure 7. Forest plot showing the effectiveness of augmented reality on skills [25,27,28,33]. Weights are from the random-effects model. RR: risk ratio.



Confidence

In total, 4 studies recorded confidence outcomes [27,29,31,35]. The total number of participants who measured their confidence was 77. There was a statistical relationship between AR training and participants' confidence, as the P value was significant according to the plot (Z=2.363; P=.02; Figure 8).

Figure 8. Forest plot showing the effectiveness of augmented reality on confidence [27,29,31,35]. Weights are from the random-effects model. RR: risk ratio.

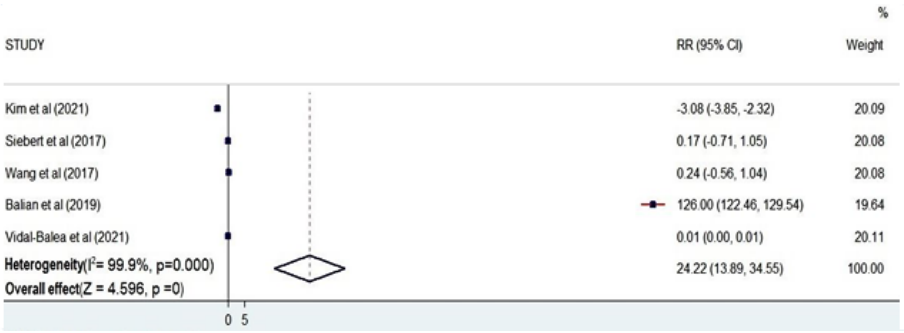


Performance Time

Performance time was assessed as an outcome measure in 5 studies [24,25,29,34,36]. The total number of participants that participated in measuring performance time was 129. There was a statistical relationship between AR training and participants' performance time because the P value was significant according to the plot (Z=4.596; P<.001; Figure 9).

XSL•FO
RenderX

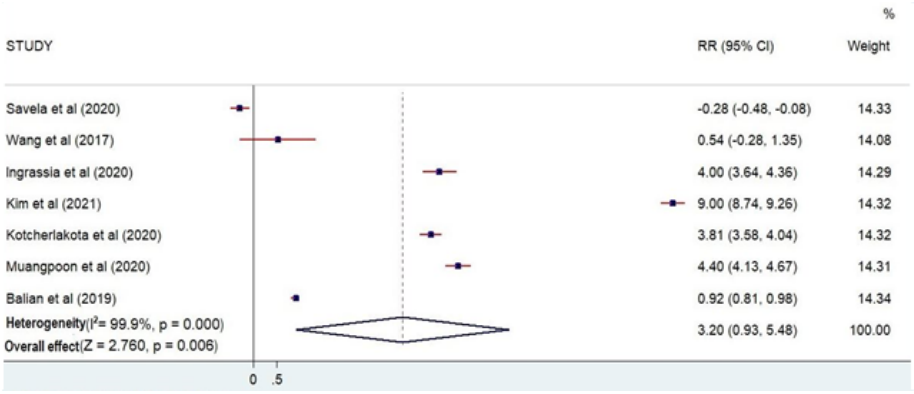
Figure 9. Forest plot showing the effectiveness of augmented reality on performance time [24,25,29,34,36]. Weights are from the random-effects model. RR: risk ratio.



Satisfaction

A total of 7 studies reported participants' satisfaction [24,25,31-35]. The total number of participants who measured their satisfaction was 543. There was a statistical relationship between AR training and participants' satisfaction according to the plot (Z=2.760; P=.006; Figure 10).

Figure 10. Forest plot showing the effectiveness of augmented reality on satisfaction [24,25,31-35]. Weights are from the random-effects model. RR: risk ratio.



Discussion

Main Findings

The usefulness of AR simulation methods in medical training was evaluated in this meta-analysis. In terms of skill outcomes, AR training did not outperform other educational methods for medical training. A study on the use of AR in medical education found that students' skills improved after the intervention [25]. In terms of skill performance scores and success rates, it was found that AR groups did not exhibit any statistical relationships, even though each of the trials that reported skills (ie, those included in our study) used different education and training methods for their control groups.

There was a statistical difference (P=.02) between the AR and control groups in terms of confidence. Participants' confidence was improved by AR when compared to that of participants in control conditions. It was found that AR training had a greater effect on participants' confidence.

The I^2 values reported throughout the *Results* section are very high, indicating that there was a huge level of heterogeneity between studies. In terms of knowledge outcomes, participants who underwent AR training did not outperform those who used other educational methods for medical training. However, a previous study that looked at how AR affects learning found that AR was more effective for knowledge retention [27] and that students are more likely to make connections between concepts when they are in an interactive learning environment. As a result, more research into the impact of AR on knowledge in medical training is needed in the future.

Between the control and AR groups, there was a significant difference in participants' satisfaction (P=.006). Of the 7 studies that recorded satisfaction as an outcome, 4 [25,32-34] indicated partial satisfaction. Some participants remarked on how difficult it is for AR technology to affect user satisfaction. As a result, we believe that participants' satisfaction with AR training may vary depending on technical factors. AR technology will be better able to satisfy users as the technology advances.

A meta-analysis of performance time was also carried out. The findings suggested that AR was more effective than other training methods at reducing performance time. However, we found that AR was more efficient than other methods in enhancing performance time. The observed heterogeneity could have been due to the various research designs and settings used in the selected studies, such as surgical projects, AR devices, and control group training methods. One study [36] on the efficiency of AR endoscopy simulation training analyzed performance time based on real-world data and found no significant difference between the AR and control groups, but the evidence quality was poor. Other studies that measured performance time as an outcome found that AR can assist operators in reducing performance time [24,25].

XSL•FO
RenderX

Strengths and Limitations

One of the main strengths of this work is being among the first meta-analyses to be conducted on the effectiveness of AR in medical training, which presents a significant contribution to the advancement of AR in the medical field. Furthermore, our work involved an assessment of 13 trials with a total of 654 participants from various countries.

Our work has some limitations as well. The first is that we only included articles that focused on AR as an intervention, which made it difficult to find a large number of studies. Second, some of the studies included in the review omitted information about allocation concealment, sequence generation, and blinding methods. Finally, the 13 trials we evaluated used different teaching and training methods for their control groups, which could lead to significant heterogeneity.

Conclusions

The effectiveness of AR training methods in the medical field was assessed in this work. The meta-analysis included 13 trials with a total of 654 participants that were completed between 2013 and 2021. In all trials, AR training was used as an intervention in AR groups, while conventional methods were used for control groups. Based on our findings, medical students' performance time, confidence, and satisfaction can be improved by using AR training and education methods. However, there was no statistical difference between the skills and knowledge that participants gained by undergoing AR training and those gained via conventional training methods. The use of AR should therefore be adopted in medical department training because of its significant effect on the performance time of participants. In general, AR should be used to expand knowledge and as a supplement to other simulation approaches to enhance clinical practice quality and safety. It is very noticeable that there are considerable gaps in the literature regarding the use of AR technology in medical training, and a limited number of studies exist, suggesting that further research efforts in this area are needed.

Acknowledgments

This research is supported by the Researchers Supporting Project (project number: RSP-2021/309), King Saud University, Riyadh, Saudi Arabia.

Authors' Contributions

YB conceptualized this meta-analysis. LFC and A Alwadain acquired funding. YB and GA conducted the investigation. YB, WNWA, GA, HA, A Alwadain, LFC, AB, and A Alghail contributed to the methodology. YB, HA, and A Alwadain were responsible for the resources used. YB was responsible for the software used. WNWA, HA, and A Alwadain supervised the meta-analysis. AB and A Alghail validated the meta-analysis. GA was responsible for data visualization. YB wrote the original manuscript draft. WNWA, GA, HA, A Alwadain, AB, LFC, and A Alghail reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 64 KB - games_v10i3e32715_app1.doc]

Multimedia Appendix 2
Search strategies.
[DOCX File , 16 KB - games_v10i3e32715_app2.docx]

Multimedia Appendix 3
Augmented reality and control groups.
[DOC File , 45 KB - games_v10i3e32715_app3.doc]

Multimedia Appendix 4
The data set.
[XLSX File (Microsoft Excel File), 13 KB - games_v10i3e32715_app4.xlsx]

Multimedia Appendix 5
Knowledge.
[XLS File (Microsoft Excel File), 27 KB - games_v10i3e32715_app5.xls]

XSL•FO
RenderX

Multimedia Appendix 6
Performance time.
[XLS File (Microsoft Excel File), 27 KB - games_v10i3e32715_app6.xls]

Multimedia Appendix 7
Satisfaction.
[XLS File (Microsoft Excel File), 28 KB - games_v10i3e32715_app7.xls]

Multimedia Appendix 8
Confidence.
[XLS File (Microsoft Excel File), 27 KB - games_v10i3e32715_app8.xls]

Multimedia Appendix 9
Skill.
[XLS File (Microsoft Excel File), 26 KB - games_v10i3e32715_app9.xls]

Multimedia Appendix 10
R and Stata codes.
[DOCX File , 13 KB - games_v10i3e32715_app10.docx]

References

1. Dickey RM, Srikishen N, Lipshultz LI, Spiess PE, Carrion RE, Hakky TS. Augmented reality assisted surgery: a urologic training tool. Asian J Androl 2016;18(5):732-734 [FREE Full text] [doi: 10.4103/1008-682X.166436] [Medline: 26620455]
2. Khor WS, Baker B, Amin K, Chan A, Patel K, Wong J. Augmented and virtual reality in surgery-the digital surgical environment: applications, limitations and legal pitfalls. Ann Transl Med 2016 Dec;4(23):454 [FREE Full text] [doi: 10.21037/atm.2016.12.23] [Medline: 28090510]
3. Zhu E, Lilienthal A, Shluzas LA, Masiello I, Zary N. Design of mobile augmented reality in health care education: A theory-driven framework. JMIR Med Educ 2015 Sep 18;1(2):e10 [FREE Full text] [doi: 10.2196/mededu.4443] [Medline: 27731839]
4. Blomqvist S, Seipel S, Engström M. Using augmented reality technology for balance training in the older adults: a feasibility pilot study. BMC Geriatr 2021 Feb 26;21(1):144 [FREE Full text] [doi: 10.1186/s12877-021-02061-9] [Medline: 33637043]
5. Malheiros O. Virtual reality and augmented reality as promoters of better health literacy: Analysing assertiveness, clear language, and positivity. In: Handbook of Research on Assertiveness, Clarity, and Positivity in Health Literacy. Hershey, Pennsylvania: IGI Global; 2022.
6. Susilawati E, Johari A, Marzal J, Anggereini E. Effectiveness of multimedia e-book and augmented reality on knowledge and skill of health science students: A systematic review. Journal of Client-Centered Nursing Care 2022 May 01;8(2):99-110 [FREE Full text] [doi: 10.32598/JCCNC.8.2.380.2]
7. Hu MH, Chiang CC, Wang ML, Wu NY, Lee PY. Clinical feasibility of the augmented reality computer-assisted spine surgery system for percutaneous vertebroplasty. Eur Spine J 2020 Jul;29(7):1590-1596. [doi: 10.1007/s00586-020-06417-4] [Medline: 32342281]
8. Zhuang Y, Sun J, Liu J. Diagnosis of chronic kidney disease by three-dimensional contrast-enhanced ultrasound combined with augmented reality medical technology. J Healthc Eng 2021 Mar 16;2021:5542822 [FREE Full text] [doi: 10.1155/2021/5542822] [Medline: 33791081]
9. Tang KS, Cheng DL, Mi E, Greenberg PB. Augmented reality in medical education: a systematic review. Can Med Educ J 2020 Mar 16;11(1):e81-e96 [FREE Full text] [doi: 10.36834/cmej.61705] [Medline: 32215146]
10. Lee LN, Kim MJ, Hwang WJ. Potential of augmented reality and virtual reality technologies to promote wellbeing in older adults. Appl Sci (Basel) 2019 Aug 30;9(17):3556 [FREE Full text] [doi: 10.3390/app9173556]
11. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: Analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
12. Juanes-Méndez JA, Marcos-Pablos S, Velasco-Marcos MA, Izard SG. Medical learning through simulations with immersive virtual reality and augmented reality devices. In: Technological Adoption and Trends in Health Sciences Teaching, Learning, and Practice. Hershey, Pennsylvania: IGI Global; 2022.
13. Kamphuis C, Barsom E, Schijven M, Christoph N. Augmented reality in medical education? Perspect Med Educ 2014 Sep;3(4):300-311 [FREE Full text] [doi: 10.1007/s40037-013-0107-7] [Medline: 24464832]
14. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020 Jan 18;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]

XSL•FO
RenderX

15. Duarte ML, Santos LR, Júnior JBG, Peccin MS. Learning anatomy by virtual reality and augmented reality. A scope review. Morphologie 2020 Dec;104(347):254-266. [doi: 10.1016/j.morpho.2020.08.004] [Medline: 32972816]
16. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Oct;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
17. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: Systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
18. Chen FQ, Leng YF, Ge JF, Wang DW, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: Meta-analysis. J Med Internet Res 2020 Sep 15;22(9):e18290 [FREE Full text] [doi: 10.2196/18290] [Medline: 32930664]
19. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
20. Williams MA, McVeigh J, Handa AI, Lee R. Augmented reality in surgical training: a systematic review. Postgrad Med J 2020 Sep;96(1139):537-542. [doi: 10.1136/postgradmedj-2020-137600] [Medline: 32229513]
21. Baashar Y, Alhussian H, Patel A, Alkawsi G, Alzahrani AI, Alfarraj O, et al. Customer relationship management systems (CRMS) in the healthcare environment: A systematic literature review. Comput Stand Interfaces 2020 Aug;71:103442 [FREE Full text] [doi: 10.1016/j.csi.2020.103442] [Medline: 34170994]
22. Moher D, Altman DG, Liberati A, Tetzlaff J. PRISMA statement. Epidemiology 2011 Jan;22(1):128; author reply 128. [doi: 10.1097/EDE.0b013e3181fe7825] [Medline: 21150360]
23. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ; Chichester, West Sussex: Wiley-Blackwell; 2019:1119536618.
24. Balian S, McGovern SK, Abella BS, Blewer AL, Leary M. Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers. Heliyon 2019 Aug 02;5(8):e02205 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e02205] [Medline: 31406943]
25. Kim SK, Lee Y, Yoon H, Choi J. Adaptation of extended reality smart glasses for core nursing skill training among undergraduate nursing students: Usability and feasibility study. J Med Internet Res 2021 Mar 02;23(3):e24313 [FREE Full text] [doi: 10.2196/24313] [Medline: 33650975]
26. Pantziaras I, Fors U, Ekblad S. Training with virtual patients in transcultural psychiatry: do the learners actually learn? J Med Internet Res 2015 Feb 16;17(2):e46 [FREE Full text] [doi: 10.2196/jmir.3497] [Medline: 25689716]
27. Albrecht UV, Folta-Schoofs K, Behrends M, von Jan U. Effects of mobile augmented reality learning compared to textbook learning on medical students: randomized controlled pilot study. J Med Internet Res 2013 Aug 20;15(8):e182 [FREE Full text] [doi: 10.2196/jmir.2497] [Medline: 23963306]
28. Noll C, von Jan U, Raap U, Albrecht UV. Mobile augmented reality as a feature for self-oriented, blended learning in medicine: Randomized controlled trial. JMIR Mhealth Uhealth 2017 Sep 14;5(9):e139 [FREE Full text] [doi: 10.2196/mhealth.7943] [Medline: 28912113]
29. Siebert JN, Ehrler F, Gervaix A, Haddad K, Lacroix L, Schrurs P, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: A randomized controlled trial. J Med Internet Res 2017 May 29;19(5):e183 [FREE Full text] [doi: 10.2196/jmir.7379] [Medline: 28554878]
30. Schiffeler N, Stehling V, Hees F, Isenhardt I. Effects of collaborative augmented reality on communication and interaction in learning contexts – Results of a qualitative pre-study. 2019 Jun 15 Presented at: 2019 ASEE Annual Conference & Exposition; June 15-19, 2019; Tampa, Florida. [doi: 10.18260/1-2--32694]
31. Kotcherlakota S, Pelish P, Hoffman K, Kupzyk K, Rejda P. Augmented reality technology as a teaching strategy for learning pediatric asthma management: Mixed methods study. JMIR Nurs 2020 Dec 02;3(1):e23963 [FREE Full text] [doi: 10.2196/23963] [Medline: 34406970]
32. Savela N, Oksanen A, Kaakinen M, Noreikis M, Xiao Y. Does augmented reality affect sociability, entertainment, and learning? A field experiment. Appl Sci (Basel) 2020 Feb 19;10(4):1392 [FREE Full text] [doi: 10.3390/app10041392]
33. Muangpoon T, Osgouei RH, Escobar-Castillejos D, Kontovounisios C, Bello F. Augmented reality system for digital rectal examination training and assessment: System validation. J Med Internet Res 2020 Aug 13;22(8):e18637 [FREE Full text] [doi: 10.2196/18637] [Medline: 32788146]
34. Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, et al. Augmented reality as a telemedicine platform for remote procedural training. Sensors (Basel) 2017 Oct 10;17(10):2294 [FREE Full text] [doi: 10.3390/s17102294] [Medline: 28994720]
35. Ingrassia PL, Mormando G, Giudici E, Strada F, Carfagna F, Lamberti F, et al. Augmented reality learning environment for basic life support and defibrillation training: Usability study. J Med Internet Res 2020 May 12;22(5):e14910 [FREE Full text] [doi: 10.2196/14910] [Medline: 32396128]
36. Vidal-Balea A, Blanco-Novoa Ó, Fraga-Lamas P, Fernández-Caramés TM. Developing the next generation of augmented reality games for pediatric healthcare: An open-source collaborative framework based on ARCore for implementing teaching, training and monitoring applications. Sensors (Basel) 2021 Mar 07;21(5):1865 [FREE Full text] [doi: 10.3390/s21051865] [Medline: 33800070]

XSL•FO
RenderX

37. Üstün U, Eryılmaz A. A research methodology to conduct effective research syntheses: Meta-analysis. Education and Science 2014;39(174):1-32 [FREE Full text] [doi: 10.15390/eb.2014.3379]
38. Üstün U. To what extent is problem-based learning effective as compared to traditional teaching in science education? A meta-analysis study. OpenMETU. 2012 Sep. URL: https://open.metu.edu.tr/bitstream/handle/11511/22032/index.pdf [accessed 2022-05-19]
39. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis, 2nd Edition. Hoboken, NJ: John Wiley & Sons; 2021.

Abbreviations

AR: augmented reality
PICO: Population, Intervention, Comparison, Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
VR: virtual reality

Edited by N Zary; submitted 06.08.21; peer-reviewed by L Suppan, P Haubruck; comments to author 12.10.21; revised version received 12.03.22; accepted 22.04.22; published 05.07.22.

Please cite as:
Baashar Y, Alkawsi G, Ahmad WNW, Alhussian H, Alwadain A, Capretz LF, Babiker A, Alghail A
Effectiveness of Using Augmented Reality for Training in the Medical Professions: Meta-analysis
JMIR Serious Games 2022;10(3):e32715
URL: https://games.jmir.org/2022/3/e32715
doi:10.2196/32715
PMID:35787488

©Yahia Baashar, Gamal Alkawsi, Wan Nooraishya Wan Ahmad, Hitham Alhussian, Ayed Alwadain, Luiz Fernando Capretz, Areej Babiker, Adnan Alghail. Originally published in JMIR Serious Games (https://games.jmir.org), 05.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effect of Virtual Reality on Pediatric Pain and Fear During Procedures Involving Needles: Systematic Review and Meta-analysis

Marta Lluesma-Vidal[1*], PhD; Raquel Carcelén González[2*], PhD; Laura García-Garcés[1], PhD; María I Sánchez-López[1], PhD; Loreto Peyro[1], PhD; Cayetana Ruiz-Zaldibar[3*], PhD

[1]Department of Nursing, Faculty of Health of Science, Universidad Cardenal Herrera-CEU, CEU Universities, Alfara del Patriarca, Valencia, Spain

[2]Department of Medicine and Surgery, Faculty of Health of Science, Universidad Cardenal Herrera-CEU, CEU Universities, Alfara del Patriarca, Valencia, Spain

[3]Department of Nursing, Faculty of Health, University of Camilo José Cela, Villanueva de la Cañada, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Raquel Carcelén González, PhD
Department of Medicine and Surgery
Faculty of Health of Science
Universidad Cardenal Herrera-CEU, CEU Universities
C/Ramon y Cajal s/n
Alfara del Patriarca, Valencia, 64115
Spain
Phone: 34 96136900 ext 64335
Email: raquel.carcelen@uchceu.es

Abstract

Background: Virtual reality (VR) is used as a distraction measure during painful clinical procedures associated with the use of needles. These procedures include vaccinations, blood draws, or the administration of medications, which can cause children to feel increased levels of pain and fear.

Objective: The objective of this study was to collect and analyze the current evidence regarding the effectiveness of VR as a tool to distract children from pain and fear during needle procedures as compared to that of standard techniques.

Methods: A systematic review and meta-analysis was performed. We included randomized clinical trials (RCTs) or quasi-RCTs with participants younger than 21 years who underwent needle procedures in which the main distraction measure used was VR and where the main outcome measure was pain. The databases searched included the PubMed, Web of Science, Scopus, PsycINFO, CINAHL, and Cochrane libraries. In this systematic review, the studies were analyzed by applying the Critical Appraisal Skills Program guide in Spanish and the Jadad scale. In the meta-analysis, the effect size of the studies was analyzed based on the results for pain and fear in children.

Results: From 665 unique search results, 21 studies were included in this systematic review, most of which reported low methodological quality. The study sample cohorts ranged from a minimum of 15 participants to a maximum of 220 participants. Ten studies were included in the meta-analysis. The global effect of using VR as a distraction measure was a significant reduction in pain (inverse variance [IV] –2.37, 95% CI –3.20 to –1.54; Z=5.58; P<.001) and fear (IV –1.26, 95% CI –1.89 to –0.63; Z=3.92; P<.001) in children in the experimental groups.

Conclusions: The quality of the studies was mostly low. The main limitations were the impossibility of blinding the participants and health care personnel to the VR intervention. Nonetheless, the use of VR as a distraction measure was effective in reducing pain and fear in children during procedures involving needles.

(JMIR Serious Games 2022;10(3):e35008) doi:10.2196/35008

KEYWORDS

virtual reality; pain; fear; pediatric; needle; child; injection; VR; systematic review; meta-analysis; paediatric

XSL•FO
RenderX

Introduction

Background

The main problems experienced in pediatric care are pain and fear. This is especially true for procedures associated with the use of needles such as vaccinations, blood draws, or the administration of medications [1,2]. This causes difficulties in the administration of health care and can result in parental dissatisfaction [3]. The International Association for the Study of Pain defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage" [4]. Pain, therefore, is a complex experience that involves sensory, cognitive, behavioral, and psychological factors [5]. In turn, fear is an immediate alarm reaction to danger, which triggers an escape behavior and an intense physiological response [6]. The pain and fear that children experience when facing needle procedures is a concern for health care professionals. Therefore, various techniques are being studied to help reduce its impact. Indeed, the administration of drugs is not always indicated to reduce pain and fear in these procedures [7]. Rather, the use of distractions during painful procedures appears to be one of the most effective ways to decrease pain and distress in children [8]. For example, music or toys have already been effectively used as distraction measures to help reduce pediatric pain. Nonetheless, virtual reality (VR) is a novel technique that has been proven to be more effective than traditional methods [3].

VR is a computer technology that creates a 3D-simulated artificial environment [5]. It usually requires wearing special glasses that cover a wide field of vision and which include motion tracking systems at the eye level [9]. These glasses can be connected to a computer or a telephone [5]. VR makes it easier to divert attention away from the painful procedure so that children will have a slower response to pain signals by counteracting them with an experience of pleasant stimuli [10,11]. Several studies have evaluated the use of VR as a distraction measure during painful procedures such as venipuncture [3,12-15], tooth extraction [16-19], or burns treatment [20-24]. However, these studies have certain limitations such as the use of small sample sizes or poor methodological quality. Comparing the findings of these studies is difficult because the works published to date have evaluated a wide breadth of invasive medical care types. Furthermore, we were able to identify only 2 systematic reviews and 1 meta-analysis that analyzed the use of VR in children. However, these studies had evaluated several medical procedures, including dental procedures, burns treatments, oncological care, or physical therapy sessions [3,25]. The variation in the procedural conditions using VR implies a lack of evidence to support its use in needle procedures. Thus, highlighting these issues, this systematic review and meta-analysis focused on the effect of VR on pain and fear during needle procedures in children.

Objectives

The general objective of this study was to collect and analyze the current evidence available regarding the effectiveness of VR as a tool to distract pediatric patients from potential pain and fear while undergoing needle procedures compared to the distractions by standard techniques. Regarding the specific objectives, our first aim was to analyze the studies included in the systematic review to assess their methodological quality. Second, our objective was to analyze the effect of the randomized controlled trials (RCTs) included in our meta-analysis.

Research Question

Is the use of VR as a distraction measure effective for reducing the perception of pain in children while performing needle procedures?

Methods

Study Design

This is a systematic review and meta-analysis of studies that evaluated the effect of VR as the main distraction measure to reduce the perception of pain in children undergoing needle procedures.

Inclusion Criteria

Studies were included in this paper based on the following criteria: (1) the participants were younger than 21 years; (2) studies where the use of VR was the primary distraction means used during needle procedures; (3) studies, including pilot studies, with an RCT or quasi-RCT methodological design; and (4) studies where the main outcome measure was pain.

Data Sources

For this study, we consulted the PubMed, Web of Science, Scopus, PsycINFO, CINAHL, and Cochrane databases. The literature search was conducted between January 2020 and June 2021. Two independent researchers comprehensively reviewed the results obtained in each of the studies and subsequently compared the selected papers.

Research Strategy

The medical subject heading keyword terms used in the search were reality, virtual, virtual reality, virtual reality headset, virtual reality exposure therapy, child*, pediatric, adolescent, intervention, program*, pain, ache, procedural, acute pain, pain perception, fear, and fears. All these terms were combined with the Boolean AND and OR functions and no filters were applied to limit the search. Search strategies were created specifically for each database by using the medical subject heading terms described above (Multimedia Appendix 1). No publication date or language restrictions were applied.

Study Selection Process

First, we evaluated the scientific literature to identify studies that met the inclusion criteria. To do this, we read the title and abstract from each of the identified papers. Two of our authors (RCG and MLV) independently performed an initial screening by reading the study titles and abstracts. After this process, the researchers discussed their results based on the predetermined inclusion and exclusion criteria. There was a 6% discrepancy in the opinions of these authors, which was resolved by further discussion to reach a consensus.

XSL•FO
RenderX

Data Extraction

Once the full-text papers were selected, 2 authors (RCG and CRZ) analyzed the studies based on their general characteristics and methodological quality. In this process, these researchers jointly extracted the relevant information from these publications. This information was transferred to 2 tables. First, the general characteristics of the studies were included in Multimedia Appendix 2. Subsequently, the methodological quality of all the studies was analyzed based on the Critical Appraisal Skills Program guide in Spanish (CASPe) scale, and this information was completed by performing a quantitative evaluation using the Jadad scale; these data are shown in Multimedia Appendix 3.

Protocol and Registration

This systematic review was registered with the Open Science Framework (Osf.io/cd8nr) in October 2021.

Data List

The general characteristics (Multimedia Appendix 2) of the studies provide information, including the following elements: author, study year and country, overall sample size, number of participants in the control and intervention groups, participant age, study type, variables and measurement instruments used, and finally, positive (P<.05), negative (P>.05), or inconclusive (±) results. Multimedia Appendix 3 provides an assessment of the methodological quality of the studies that we included in this review according to the CASPe [26]. This tool organizes data about each study into 3 sections: validity, results, and applicability. We used the Jadad scale [27], which assesses research quality on a scale of 0 to 5 points according to the responses to a series of questions, to complete this information. Scores below 3 points suggested that little methodological rigor had been applied during the study in question. This allowed us to objectively assess the following parameters: random sequence generation, allocation concealment, blinding of participants and personnel, and blinding to the outcome assessment. To guarantee the quality of this meta-analysis, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement guidelines [28] (Tables S4 and S5 of Multimedia Appendices 4 and 5, respectively).

Risk of Bias Assessment

The Cochrane Collaboration Risk of Bias Tool [29] was used to assess the risk of bias in the studies included in the meta-analysis in 5 categories: selection bias, performance bias, detection bias, attrition bias, and reporting bias. For selection bias, which refers to the introduction of differences between groups at baseline, random sequence generation and allocation concealment were judged. Performance bias was analyzed based on blinding of the participants and personnel. Detection bias referred to blinding of the outcome assessors. Attrition bias included different rates of withdrawals between groups and was judged according to the proportion of incomplete outcome data. Finally, reporting bias described selective reporting.

The Cochrane Collaboration Handbook for Systematic Reviews for Interventions was used to analyze the risk of bias from studies not included in the meta-analysis. This analysis included selection bias when randomization was analyzed, performance bias when blinding between participants and personnel was tested, detection bias when blinding between participants and outcome assessors was tested, attrition bias where dropouts were analyzed, and reporting bias where they were analyzed, and the outcomes were selectively reported [29].

Analysis of the Meta-analysis Data

Employing the random effects model in Review Manager software (RevMan v.5.2; Cochrane Collaboration), 2 meta-analyses were carried out to examine the overall effect of the intervention on pain and fear in children. We used this model because we wanted to limit overestimation of the effect size. The studies included had an RCT design and contained complete statistical information; the effects were expressed as mean differences with a 95% CI. The heterogeneity of the studies was assessed by calculating the I^2 statistic, and the variance between the studies was examined by calculating Tau^2. When the significance level was set at .05, the heterogeneity of the studies we included was high for both these variables (94% and 96%, respectively; P<.01). Lastly, to increase the precision of the effect size estimator, the effect sizes proposed by Cohen [30] were calculated (small effect, d=0.20; medium effect, d=0.50; and large effect, d=0.80).

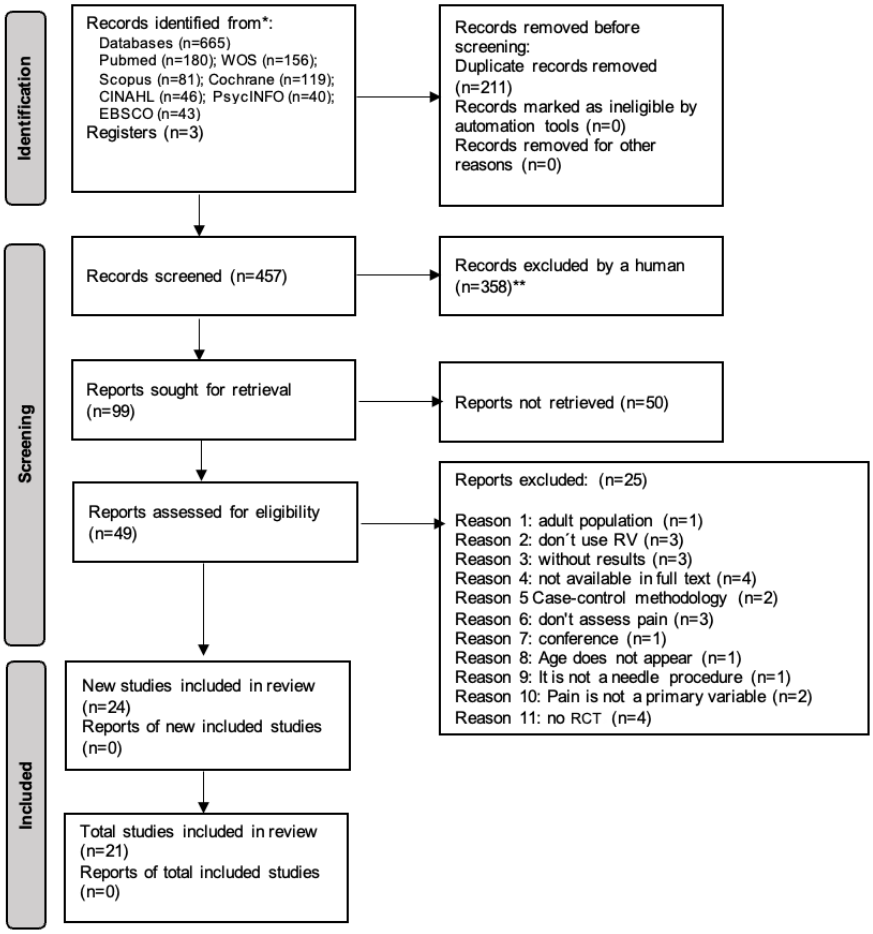
Results

Search Results

As shown in Figure 1, our initial search returned a total of 665 papers. After eliminating 211 duplicates, 2 researchers (RCG and MLV) initially screened the 454 studies by reading their titles and abstracts. There was a 6% discrepancy in their opinions, which was resolved by reaching a consensus based on the eligibility criteria of the papers. This selection further reduced the sample to 96 manuscripts. Reading the full texts of these papers revealed that only 46 papers had focused on the use of VR to reduce pain during procedures involving needles, some of which had also addressed fear in these patients. Lastly, 3 of our authors (RCG, MLV, and CRZ) critically read all these papers and excluded another 25 papers because they did not meet the inclusion criteria, as described in Figure 1. Thus, 21 studies were finally included in this systematic review, and only 10 were eligible for inclusion in the meta-analysis [31] (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart showing the screening and selection process for the papers included in this systematic review and meta-analysis. Created using the guidelines on Page et al [31]. WOS: Web of Science; *Consider if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers); **If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.



General Characteristics of the Studies

Multimedia Appendix 2 summarizes the main characteristics and results of the selected studies. The 21 studies ranged from the year 2002 to 2021; 76% (16/21) of them had been published between 2018 and 2021 [32-47], while the rest had been published between 2002 and 2007 [48-52]. Most of the research (11/21, 52%) had been conducted in North America [33,35-37,39,40,48-52], but 14% (3/21) of the work was from Europe [34,35,37] and 33% (7/21) had been performed in Asia [32,36,38-40,44,47]. Regarding the contexts of these VR studies, 95% (20/21) of them had been carried out in hospitals [32-50,52], while 5% (1/21) had been completed in primary health care centers [51]. Most of the procedures involving needles in which VR had been used were venipunctures (14/21, 67%) [32,34,35,37-43,45,47,49,51], followed by subcutaneous venous puncture for reservoir-type venous access (4/21, 19%) [33,36,48,52]. One study had used VR during lumbar punctures [50]. The remaining one had conducted research analyzing venipunctures or intramuscular injections [42]. The study sample cohorts ranged from a minimum of 15 [41] to a maximum of 220 individuals [33]; 43% (9/21) of the studies had analyzed a sample comprising fewer than 100 participants [35,37,41-43,48,49,51,52]. In the different studies, the age of the children ranged from 4 years to 21 years, while the study duration varied between 14 weeks and 20 months; 57% (12/21) of the studies collected data for less than a year [32,34-37,39-41,43,44,46,47], 10% (2/21) had done so for 13 months or more [33,45], and 33% (7/21) of them had not reported this information [38,42,48-52]. We identified most of the studies (16/21, 76%) as RCTs [32-34,36-40,42-47,49,52] but 10% (2/21) were pilot studies [48,50] and 10% (2/21) were quasi-experimental studies [35,41].

Quality of the Studies

We assessed the quality of the studies according to the CASPe and Jadad guidelines (see Multimedia Appendix 3). Only 14% (3/21) of the studies were rated as high quality [37,44,45], with the remaining 86% (18/21) being rated as low quality [32-36,38-43,46-52]. Specifically regarding the random assignment of patients to the study groups, 10% (2/21) [34,35] of the studies had not carried out randomization. The participants had been randomly assigned in 19% (4/21) of the studies [32,49,50,52] but none of these authors had specified the type of procedure they had used to perform the randomization, and this information was uncertain in another study (1/21, 5%) [41]. The participants had been randomly assigned in the remaining 71% (15/21) of the studies [33,36-40,42-48,51]. Given the active nature of these interventions, most of the studies had not blinded the participants to their group assignment. Moreover, only 4 (19%) of them [37,39,45,52] had blinded the group assignment to the observers or health professionals, although none of them

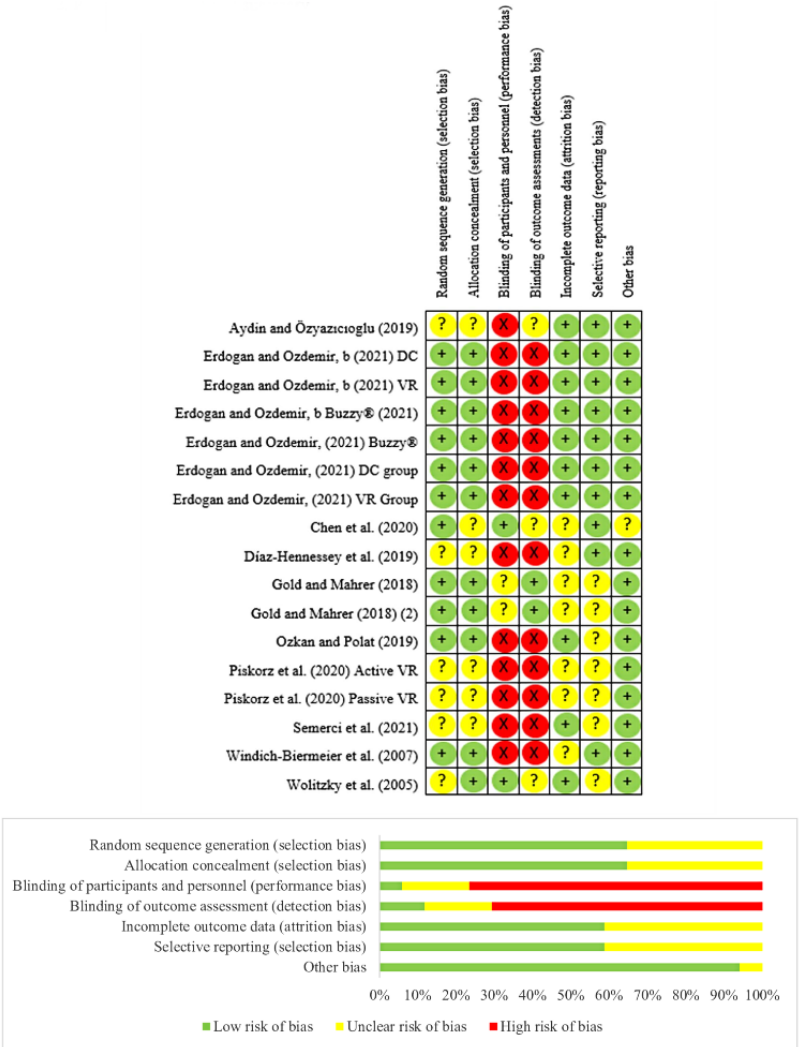
XSL•FO
RenderX

had been able to maintain this blinding until the end of the work. The preintervention VR and control group characteristics were similar in terms of sex, age, and other sociodemographic variables in 15 (71%) of the 21 studies [32,33,36-39,42-45,47-49,51,52]. In 91% (19/21) of the cases, both groups had been treated in a similar way, regardless of the intervention that had been performed [32,33,35-39,41-45,47-53]. There were insufficient reports on the flow of participants through the studies in 38% (8/21) of the papers retrieved, which made it difficult to determine the level of dropouts [34,41,42,48-52]. Only 1 study (5%) provided information about the effect size [33]. The cohorts comprised 15 to 59 children in 43% (9/21) of the studies, and the authors themselves classified these samples as small [35,41-43,48-52]. Furthermore, 5% (1/21) of the samples were of children with specific pathologies [34]. Regarding extrapolation of the results, the data could only be generalized or considered for extrapolation in 8 of the 21 papers we reviewed [32,33,37,39,40,43-45]. Additionally, only 43% (9/21) of the studies had collected information about the participant acceptance and satisfaction with the VR intervention [32,35,39-41,44,46-48]. Finally, the benefits of the intervention had exceeded the costs or damages that could have been produced in all of the cases [32-52].

Risk of Bias

The Cochrane Collaboration Risk of Bias Tool [29] was used to assess the risk of bias of the 10 studies included in the metanalysis by 2 reviewers. Based on these tools, only 1 of the studies was at high risk of bias, 8 at unclear risk of bias, and 1 at low risk of bias (Figure 2). Based on the Cochrane Collaboration criteria for different types of bias, we analyzed the 11 studies not included in the meta-analysis. As shown in Multimedia Appendix 3, the biases related to blinding, both of the participants of the personnel as well as to the outcome assessment, reached the highest levels in 82% (9/11) of the studies. Of the 11 studies, most of the studies had a moderate risk of bias (5/11, 46%); 3 (27%) studies were identified as having a high risk of bias and 2 (18%) studies had a low risk of bias. One study (9%) was classified as having a low risk of bias but no information on blinding could be obtained.

Figure 2. Risk of bias graph and summary [34,36,40,41,43,45,47,51,52].



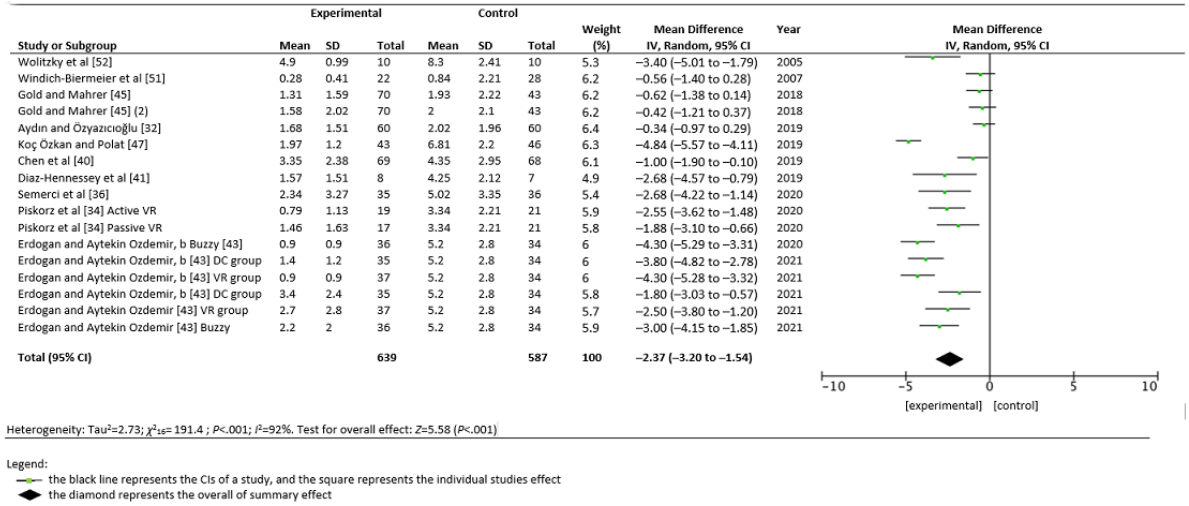
Effects of VR on the Perception of Pain

The studies were heterogeneous in both the measured outcomes (I^2=89-92). We were able to analyze the effect size of the pain studies in 10 of the 21 studies (Figure 3). The main results showed statistically significant differences in favor of the experimental group in the studies by Wolitzky et al [52] (d=1.85; inverse variance [IV] –3.40, 95% CI –5.01 to –1.79) and

XSL•FO
RenderX

Diaz-Hennessey et al [41] (*d*=1.43; IV –2.68, 95% CI –4.57 to –0.79). Likewise, pain was significantly reduced in the studies by Koç Özkan and Polat [47] (*d*=0.17; IV –4.84, 95% CI –5.57 to –4.11), the intervention by Piskorz et al [34] using both passive VR (*d*=0.97; IV –1.88, 95% CI –3.10 to –0.66) and active VR (*d*=1.45; IV –2.55, 95% CI –3.62 to –1.48), and in the studies by Erdogan and Aytekin Ozdemir [43] in VR versus a control group (*d*=0.89; IV –2.5, 95% CI –3.80 to –1.20). The study by Chen et al [40] also found a significant reduction in pain in the intervention group (*d*=0.37; IV –1.00, 95% CI –1.90 to –0.10). As shown in Figure 3, the global effect of using VR as a distraction measure had significantly reduced pain in children in the experimental groups (IV –2.37, 95% CI –3.20 to –1.54; *Z*=5.58; *P*<.001).

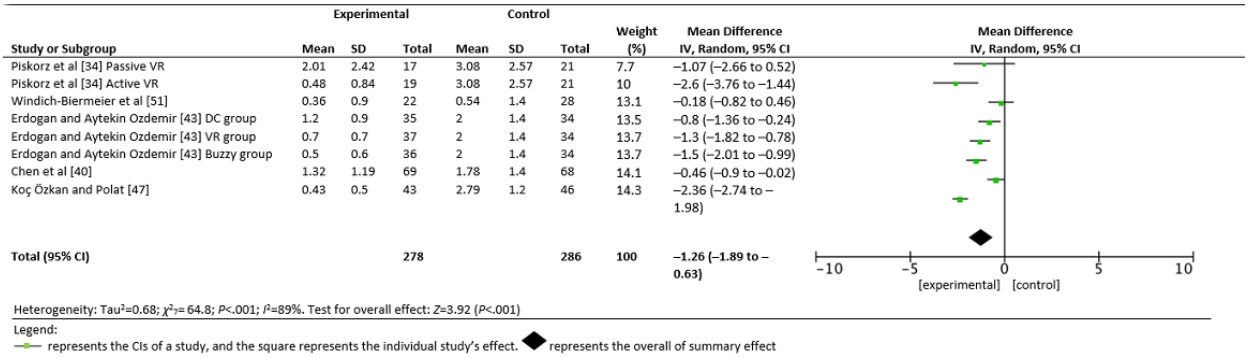
Figure 3. A random forest plot of the association between pain and study group (control vs virtual reality) [34,36,40,41,43,45,47,51,52]. b: Wong-Baker Faces Pain Rating Scale; Buzzy: a device that applies local cold and vibration at the injection site; DC: distraction card; IV: inverse variance; VR: virtual reality.



Effects of VR on Fear

We were only able to analyze the fear variable in 5 of the 21 studies. The use of VR produced a statistically significant reduction in fear in the experimental groups in the study by Chen et al [40] (*d*=0.35; IV –0.46, 95% CI –0.90 to –0.02) and a large reduction in the Koç Özkan and Polat study [47] (*d*=0.17; IV –2.36, 95% CI –2.74 to –1.98). Likewise, fear was significantly reduced in the studies by Erdogan and Aytekin Ozdemir [43] in the VR versus control group (*d*=1.17; IV –1.30, 95% CI –1.82 to –0.78) and the intervention by Piskorz et al [34] in active VR (*d*=1.36; IV –2.60, 95% CI –3.76 to –1.44]. As shown in Figure 4, the global effect of using VR as a distraction measure had significantly reduced the perception of fear in children in the experimental groups (IV –1.26, 95% CI –1.89 to –0.63; *Z*=3.92; *P*<.001).

Figure 4. A random forest plot of the association between fear and study group (control vs virtual reality) [34,40,43,47,51]. Buzzy: a device that applies local cold and vibration at the injection site; DC: distraction card; IV: inverse variance; VR: virtual reality.



Discussion

To the best of our knowledge, this is the first systematic review with a meta-analysis designed to examine the effectiveness of the use of VR as a distraction measure to reduce pain and fear in the pediatric population during procedures involving needles. Based on the high effect sizes that we found, our results suggest that VR distraction is possibly more effective than the habitual routine or other distractions used during needle procedures to reduce the perception of pain and fear felt by children. It is difficult to compare these results with those of other studies because most of them included different medical processes or did not analyze the effect on the children's fear. However, other meta-analyses found similar results, indicating that the effects of VR are beneficial in reducing fear during medical processes involving pain, especially in children [54]. However, these comparisons must be analyzed with caution because neither the studies included nor their participants were homogeneous in terms of age or characteristics, the medical procedures analyzed, or the tools used to measure pain.

XSL•FO
RenderX

Most of the papers included in this review found that VR had a positive effect by helping to reduce pain in children. Of note, all the studies that had included more than 100 participants and had used the Wong-Baker Faces Pain Rating Scale (WBFPS) had reported statistically significant results. This may be because this visual assessment scale is more effective in assessing pain in children than other scales that use numerical assessment scales such as the visual analog scale (VAS) for pain [55]. Although the VAS is a reliable method for assessing acute pain, children younger than 7 years may have difficulty in its use, as indicated by the reduced reliability of the results reported in these studies [56]. In addition, the VAS and WBFPS have been widely used in studies evaluating pain in other procedures such as wound healing [57], physiotherapy sessions after complex surgical interventions [58], or dental procedures [59] in which they produced positive results.

Most of the papers included in this review [32,33,35,37-41,44,45,47-52] had analyzed the effect of VR on pain and fear in pediatric patients with cancer during venipuncture or reservoir puncture procedures. Furthermore, most of the studies we retrieved (20/21, 95%) had been carried out in hospitals, while only 5% (1/21) had been carried out in primary health care centers. This may have been a result of the health care provision resources available at the sites where these previous studies had been carried out, given that most of this work had been carried out in hospitals, thanks to the teaching function of these centers [60-62]. These data indicate that scant research has been carried out for this level of care, which is surprising, considering that needle procedures are frequent in primary care contexts because of the systematic vaccination programs carried out in the pediatric population. Among other possible explanations, perhaps this lack of research can be explained by health care staff overload or low levels of motivation among professionals or toward the support of research [63-67]. However, 2 study protocols have recently been published that will aim to evaluate the effectiveness of VR against pain during vaccination in the pediatric population through RCTs with estimated sample sizes of 100 [68] to more than 400 participants [69].

Although we found that VR is effective in reducing children's fear, very few studies have demonstrated the usefulness of VR in reducing fear during procedures involving needles [40,47]. Thus, the absence of a validated scale to measure this variable may be inhibiting its proper evaluation [70]. According to Taddio et al [71], most studies that measure fear do so by using questionnaires developed by the investigators, nonvalidated scales, or scales for measuring anxiety [72,73]. Thus, this review reveals the lack of consensus on the most appropriate instruments for evaluating and clearly differentiating between fear and distress in the pediatric population. Although in clinical practice, the difference between fear, anxiety, and stress may not always be relevant, these represent different theoretical constructs, which are not always rigorously differentiated. Notwithstanding, both fear and distress are important factors that are related to and impact the pain perceived by children [74,75].

Of note, the quality of the studies included in this systematic review (based on CASPe and Jadad assessments) was mostly low. However, some studies with low quality or even small samples showed important effects. We assume that in the future, a meta-regression model could be used to expand existing knowledge about these intervention types and their methodological quality. For this reason, this systematic review and meta-analysis highlights the need to design and implement new research with high methodological quality that would allow extraneous variables to be isolated, favoring the cause-effect relationship. The principal reasons for the studies included in this meta-analysis to be of low quality were that it was nearly impossible to blind both the participants and health care personnel to the VR intervention because of the nature of these devices [76]. Furthermore, in many cases, the absence of randomization was justified for ethical reasons. Indeed, more than half of the studies we examined had considered small sample sizes of fewer than 100 participants [77], which, in addition to being unreliable and inefficient, can lead to overestimation of the study effect size and can produce low reproducibility of the results. Finally, chronological age and neurological development are among the factors that influenced children's perceptions of pain and fear of procedures involving needles, and therefore, adjusting the age of children to less than 21 years should be considered in future studies [78]. Blinding and randomization are also the issues that were identified in the risk of bias analysis of studies not included in the meta-analysis. The studies included in the meta-analysis generally had a low level of risk, while studies not included tended to have a higher level of risk of bias. This may be due both to the fact that meta-analysis studies are more robust and to the use of different measurement tools in these papers.

The main limitations of this work were, on the one hand, the lack of studies with nonsignificant results available in the scientific literature. This meant that we may not have included all the relevant studies, and therefore, it was not possible to control for publication bias [79]. On the other hand, although the random effects model that we used favored the most realistic observation of the data by specifically weighting each study, the heterogeneity of the included studies, both in terms of their outcome measures and their methodological approaches, means that we must be cautious about the interpretation of our results. This problem was also identified in a similar recent meta-analysis in which heterogeneity was found in studies with young patients [54]. Finally, the studies included did not address the effect of VR in children younger than 4 years, which implies a limitation of the results when it comes to generalizing this effect in all children. Based on all the above, the methodological design of future work must adequately calculate the required sample sizes and use appropriate sampling, participant study group allocations, and blinding techniques to be able to extrapolate any data obtained to the wider pediatric population. This review was limited by the quality of the studies it included. Generalization of these findings to younger children should also be done with caution because the studies we considered had not included children younger than 4 years.

In conclusion, the findings of this review indicate that VR could be a feasible distraction measure to reduce the perception of pain and fear in the pediatric population during procedures involving needles. However, these results are limited by the

XSL•FO
RenderX

heterogeneity of the studies included. In this sense, more trials with larger sample sizes and quality methodological techniques will be needed in the future.

Acknowledgments

This work was supported by a grant from the University CEU Cardenal Herrera (ICLINIC/1906).

Authors' Contributions

MLV and CRZ conceptualized and designed the study, drafted the initial manuscript, designed the data collection instruments, and reviewed the manuscript. RCG collected the data, carried out the analysis, and revised the manuscript. LP drafted the initial manuscript. LGG and MISL critically reviewed the manuscript for important intellectual content. All the authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOC File , 33 KB - games_v10i3e35008_app1.doc]

Multimedia Appendix 2
General characteristics and results of the studies in this review.
[DOC File , 211 KB - games_v10i3e35008_app2.doc]

Multimedia Appendix 3
Evaluation of the methodological quality.
[DOC File , 77 KB - games_v10i3e35008_app3.doc]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[DOC File , 88 KB - games_v10i3e35008_app4.doc]

Multimedia Appendix 5
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis)_2020_Abstract_checklist.
[DOC File , 48 KB - games_v10i3e35008_app5.doc]

References

1. Hoffman HG. Virtual-reality therapy. Sci Am 2004 Aug;291(2):58-65. [doi: 10.1038/scientificamerican0804-58] [Medline: 15298120]
2. Hedén L, von Essen L, Ljungman G. Children's self-reports of fear and pain levels during needle procedures. Nurs Open 2020 Jan;7(1):376-382 [FREE Full text] [doi: 10.1002/nop2.399] [Medline: 31871722]
3. Eijlers R, Utens EMWJ, Staals LM, de Nijs PFA, Berghmans JM, Wijnen RMH, et al. Systematic Review and Meta-analysis of Virtual Reality in Pediatrics. Anesthesia & Analgesia 2019;129(5):1344-1353. [doi: 10.1213/ane.0000000000004165] [Medline: 31136330]
4. No authors listed. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms. Prepared by the International Association for the Study of Pain, Subcommittee on Taxonomy. Pain Suppl 1986;3:S1-226. [Medline: 3461421]
5. Arane K, Behboudi A, Goldman RD. Virtual reality for pain and anxiety management in children. Canadian Family Physician 2017 Dec;63(12):932-934. [Medline: 29237632]
6. McMurtry CM, Riddell RP, Taddio A, Racine N, Asmundson GJG, Noel M, et al. Far from "Just a poke": Common painful needle procedures and the development of needle fear. Clin J Pain 2015;31(10):S3-11. [doi: 10.1097/ajp.0000000000000272] [Medline: 26352920]
7. Gates M, Hartling L, Shulhan-Kilroy J, MacGregor T, Guitard S, Wingert A, et al. Digital technology distraction for acute pain in children: A Meta-analysis. Pediatrics 2020;145(2):e20191139. [doi: 10.1542/peds.2019-1139] [Medline: 29237632]

XSL•FO
RenderX

8. Birnie KA, Noel M, Parker JA, Chambers CT, Uman LS, Kisely SR, et al. Systematic review and meta-analysis of distraction and hypnosis for needle-related pain and distress in children and adolescents. J Pediatr Psychol 2014 Sep;39(8):783-808 [FREE Full text] [doi: 10.1093/jpepsy/jsu029] [Medline: 24891439]
9. Wismeijer A, Vingerhoets AJJM. The use of virtual reality and audiovisual eyeglass systems as adjunct analgesic techniques: A review of the literature. Annals of Behavioral Medicine: a publication of the Society of Behavioral Medicine 2005;30(3):268-278. [doi: 10.1207/s15324796abm3003_11] [Medline: 16336078]
10. Schneider SM, Workman ML. Effects of virtual reality on symptom distress in children receiving chemotherapy. Cyberpsychol Behav 1999;2(2):125-134. [doi: 10.1089/cpb.1999.2.125] [Medline: 19178248]
11. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
12. Dunn A, Patterson J, Biega CF, Grishchenko A, Luna J, Stanek JR, et al. A Novel Clinician-Orchestrated Virtual Reality Platform for Distraction During Pediatric Intravenous Procedures in Children With Hemophilia: Randomized Controlled Trial. JMIR Serious Games 2019 Jan 09;7(1):e10902 [FREE Full text] [doi: 10.2196/10902] [Medline: 30626567]
13. Gold JI, Belmont KA, Thomas DA. The neurobiology of virtual reality pain attenuation. Cyberpsychol Behav 2007 Aug;10(4):536-544. [doi: 10.1089/cpb.2007.9993] [Medline: 17711362]
14. Miller K, Tan X, Hobson AD, Khan A, Ziviani J, O Brien E, et al. A Prospective Randomized Controlled Trial of Nonpharmacological Pain Management During Intravenous Cannulation in a Pediatric Emergency Department. Pediatric Emergency Care 2016 Jul;32(7):444-451. [doi: 10.1097/pec.0000000000000778] [Medline: 27380603]
15. Nasirzadeh A, Farzin SF, Ghodrati M, Shareinia H. Comparing the Effect of Guided Visualization and Virtual Reality Techniques on Cannulation Pain in Hemodialysis Patients. Journal of Research in Medical and Dental Science 2019;7(3):55-62 [FREE Full text]
16. Atzori B, Lauro Grotto RL, Giugni A, Calabrò M, Alhalabi W, Hoffman HG. Virtual Reality Analgesia for Pediatric Dental Patients. Front Psychol 2018 Nov;9:2265 [FREE Full text] [doi: 10.3389/fpsyg.2018.02265] [Medline: 30532720]
17. Baschirotto N, Dos FDS, Gonzalez M, Pereira V, Leao M. Effectiveness of Virtual Reality Glasses as a Distraction for Children During Dental Care. Pediatric Dentistry 2020 Mar;42(2):93-102. [Medline: 32276674]
18. Niharika P, Reddy N, Srujana P, Srikanth K, Daneswari V, Geetha K. Effects of distraction using virtual reality technology on pain perception and anxiety levels in children during pulp therapy of primary molars. J Indian Soc Pedod Prev Dent 2018;36(4):364-369. [doi: 10.4103/jisppd.jisppd_1158_17] [Medline: 30324926]
19. Nunna M, Dasaraju RK, Kamatham R, Mallineni SK, Nuvvula S. Comparative evaluation of virtual reality distraction and counter-stimulation on dental anxiety and pain perception in children. J Dent Anesth Pain Med 2019 Oct;19(5):277-288 [FREE Full text] [doi: 10.17245/jdapm.2019.19.5.277] [Medline: 31723668]
20. Schmitt YS, Hoffman HG, Blough DK, Patterson DR, Jensen MP, Soltani M, et al. A randomized, controlled trial of immersive virtual reality analgesia, during physical therapy for pediatric burns. Burns 2011 Feb;37(1):61-68 [FREE Full text] [doi: 10.1016/j.burns.2010.07.007] [Medline: 20692769]
21. Miller K, Rodger S, Kipping B, Kimble R. A novel technology approach to pain management in children with burns: A prospective randomized controlled trial. Burns 2011 May;37(3):395-405. [doi: 10.1016/j.burns.2010.12.008] [Medline: 21306828]
22. Khadra C, Ballard A, Déry J, Paquin D, Fortin J, Perreault I, et al. Projector-based virtual reality dome environment for procedural pain and anxiety in young children with burn injuries: a pilot study. JPR 2018 Feb;Volume 11:343-353. [doi: 10.2147/jpr.s151084] [Medline: 29491717]
23. Ford CG, Manegold EM, Randall CL, Aballay AM, Duncan CL. Assessing the feasibility of implementing low-cost virtual reality therapy during routine burn care. Burns 2018 Jun;44(4):886-895 [FREE Full text] [doi: 10.1016/j.burns.2017.11.020] [Medline: 29305105]
24. Harper H, Hirt P, Lev-Tov H. The use of virtual reality in non-burn dermatological care - a review of the literature. J Dermatolog Treat 2022 Feb;33(1):48-53. [doi: 10.1080/09546634.2020.1743812] [Medline: 32174192]
25. Lambert V, Boylan P, Boran L, Hicks P, Kirubakaran R, Devane D. Virtual reality distraction for acute pain in children. Cochrane database Syst Rev Internet 2020 Oct;10:CD010686 [FREE Full text] [doi: 10.1002/14651858.CD010686.pub2] [Medline: 33089901]
26. Cabello JB. Plantilla para ayudarte a entender un Ensayo Clínico. Guías CASPe de Lectura Crítica de la Literatura Médica. Alicante. 2005. URL: https://tinyurl.com/2p9efj7k [accessed 2022-07-21]
27. Armijo OS, Gazzi ML, Gadotti I, Fuentes J, Stanton T, Magee D. Scales to Assess the Quality of Randomized Controlled Trials: A Systematic Review. Physical Therapy 2008 Feb;88(2):156-175. [doi: 10.2522/ptj.20070147] [Medline: 18073267]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
29. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. United Kingdom: John Wiley & Sons, Ltd; 2011:197-255.
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press; 1988.

XSL•FO
RenderX

31. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
32. Aydın, Özyazıcıoğlu N. Using a Virtual Reality Headset to Decrease Pain Felt During a Venipuncture Procedure in Children. J Perianesth Nurs 2019 Dec;34(6):1215-1221. [doi: 10.1016/j.jopan.2019.05.134] [Medline: 31495558]
33. Caruso TJ, George A, Menendez M, De Souza E, Khoury M, Kist MN, et al. Virtual reality during pediatric vascular access: A pragmatic, prospective randomized, controlled trial. Paediatr Anaesth 2020 Feb;30(2):116-123. [doi: 10.1111/pan.13778] [Medline: 31785015]
34. Piskorz J, Czub M, Sulzickaja B, Kilis-Pstrusinka K. Mobile virtual reality distraction reduces needle pain and stress in children. J Psychosoc Res Cybersp. URL: https://cyberpsychology.eu/article/view/11547/11316 [accessed 2021-04-13]
35. Piskorz J, Czub M. Effectiveness of a virtual reality intervention to minimize pediatric stress and pain intensity during venipuncture. J Spec Pediatr Nurs 2017 Nov 20;23(1):e12201. [doi: 10.1111/jspn.12201] [Medline: 29155488]
36. Semerci R, Akgün Kostak M, Eren T, Avci G. Effects of Virtual Reality on Pain During Venous Port Access in Pediatric Oncology Patients: A Randomized Controlled Study. J Pediatr Oncol Nurs 2021;38(2):142-151. [doi: 10.1177/1043454220975702] [Medline: 33300448]
37. Walther-Larsen S, Petersen T, Friis S, Aagaard G, Drivenes B, Opstrup P. Immersive Virtual Reality for Pediatric Procedural Pain: A Randomized Clinical Trial. Hosp Pediatr 2019 Jul;9(7):501-507. [doi: 10.1542/hpeds.2018-0249] [Medline: 31160472]
38. Wong CL, Li CK, Chan CWH, Choi KC, Chen J, Yeung MT, et al. Virtual Reality Intervention Targeting Pain and Anxiety Among Pediatric Cancer Patients Undergoing Peripheral Intravenous Cannulation. Cancer Nurs 2020 Jun 4;44(6):435-442. [doi: 10.1097/ncc.0000000000000844] [Medline: 32511154]
39. Gerçeker GÖ, Binay Ş, Bilsin E, Kahraman A, Yılmaz HB. Effects of Virtual Reality and External Cold and Vibration on Pain in 7- to 12-Year-Old Children During Phlebotomy: A Randomized Controlled Trial. J Perianesth Nurs 2018 Dec;33(6):981-989. [doi: 10.1016/j.jopan.2017.12.010] [Medline: 29559294]
40. Chen Y, Cheng S, Lee P, Lai C, Hou I, Chen C. Distraction using virtual reality for children during intravenous injections in an emergency department: A randomised trial. J Clin Nurs 2020 Feb;29(3-4):503-510. [doi: 10.1111/jocn.15088] [Medline: 31715039]
41. Diaz-Hennessey S, O'Shea ER, King K. Virtual reality: augmenting the acute pain experience in children. Pediatric Nursing. 2019. URL: https://www.thefreelibrary.com/Virtual+Reality%3A+Augmenting+the+Acute+Pain+Experience+in+Children.-a0592139557 [accessed 2022-07-21]
42. Dumoulin S, Bouchard S, Ellis J, Lavoie KL, Vézina MP, Charbonneau P, et al. A Randomized Controlled Trial on the Use of Virtual Reality for Needle-Related Procedures in Children and Adolescents in the Emergency Department. Games Health J 2019 Aug;8(4):285-293. [doi: 10.1089/g4h.2018.0111] [Medline: 31135178]
43. Erdogan B, Aytekin Ozdemir A. The Effect of Three Different Methods on Venipuncture Pain and Anxiety in Children: Distraction Cards, Virtual Reality, and Buzzy® (Randomized Controlled Trial). J Pediatr Nurs 2021;58:e54-e62. [doi: 10.1016/j.pedn.2021.01.001] [Medline: 33485746]
44. Özalp Gerçeker GÖ, Ayar D, Özdemir EZ, Bektaş M. Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5-12 years old: A randomised controlled study. J Clin Nurs 2020 Apr;29(7-8):1151-1161. [doi: 10.1111/jocn.15173] [Medline: 31889358]
45. Gold J, Mahrer N. Is Virtual Reality Ready for Prime Time in the Medical Space? A Randomized Control Trial of Pediatric Virtual Reality for Acute Procedural Pain Management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
46. Goldman RD, Behboudi A. Virtual reality for intravenous placement in the emergency department-a randomized controlled trial. Eur J Pediatr 2021 Mar;180(3):725-731. [doi: 10.1007/s00431-020-03771-9] [Medline: 32779029]
47. Koç Özkan T, Polat F. The Effect of Virtual Reality and Kaleidoscope on Pain and Anxiety Levels During Venipuncture in Children. J Perianesth Nurs 2020 Apr;35(2):206-211. [doi: 10.1016/j.jopan.2019.08.010] [Medline: 31759833]
48. Gershon J, Zimand E, Pickering M, Rothbaum BO, Hodges L. A pilot and feasibility study of virtual reality as a distraction for children with cancer. J Am Acad Child Adolesc Psychiatry 2004 Oct;43(10):1243-1249. [doi: 10.1097/01.chi.0000135621.23145.05] [Medline: 15381891]
49. Gold JI, Kim SH, Kant AJ, Joseph MH, Rizzo A. Effectiveness of virtual reality for pediatric pain distraction during I.V. placement. Cyberpsychol Behav 2006 Apr;9(2):207-212. [doi: 10.1089/cpb.2006.9.207] [Medline: 16640481]
50. Wint SS, Eshelman D, Steele J, Guzzetta CE. Effects of Distraction Using Virtual Reality Glasses During Lumbar Punctures in Adolescents With Cancer. Oncology Nursing Forum 2007 Feb 14;29(1):E8-E15. [doi: 10.1188/02.onf.e8-e15]
51. Windich-Biermeier A, Sjoberg I, Dale JC, Eshelman D, Guzzetta CE. Effects of distraction on pain, fear, and distress during venous port access and venipuncture in children and adolescents with cancer. J Pediatr Oncol Nurs 2007;24(1):8-19. [doi: 10.1177/1043454206296018] [Medline: 17185397]
52. Wolitzky K, Fivush R, Zimand E, Hodges L, Rothbaum BO. Effectiveness of virtual reality distraction during a painful medical procedure in pediatric oncology patients. Psychology & Health 2005 Dec;20(6):817-824. [doi: 10.1080/14768320500143339]

XSL•FO
RenderX

53. Goldman RD, Behboudi A. Virtual reality for intravenous placement in the emergency department-a randomized controlled trial. Eur J Pediatr 2021 Mar;180(3):725-731. [doi: 10.1007/s00431-020-03771-9] [Medline: 32779029]
54. Huang Q, Lin J, Han R, Peng C, Huang A. Using Virtual Reality Exposure Therapy in Pain Management: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Value Health 2022 Feb;25(2):288-301. [doi: 10.1016/j.jval.2021.04.1285] [Medline: 35094802]
55. Garra G, Singer A, Taira B, Chohan J, Cardoz H, Chisena E. Validation of the Wong-Baker FACES pain rating scale in pediatric emergency department patients. Academic Emergency Medicine: oficial journal of the Society for the Academic Emergency Medicine 2010 Jan;17(1):50-54. [doi: 10.1111/j.1553-2712.2009.00620.x] [Medline: 20003121]
56. Bárcena Fernández E. Manejo del dolor pediátrico en el centro de salud. Rev Pediatr Aten Primaria 2014 Jun;16:37-43. [doi: 10.4321/s1139-76322014000200005]
57. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The Effect of Virtual Reality Distraction on Pain Relief During Dressing Changes in Children with Chronic Wounds on Lower Limbs. Pain Manag Nurs 2015 Oct;16(5):685-691. [doi: 10.1016/j.pmn.2015.03.001] [Medline: 25972074]
58. Steele E, Grimmer K, Thomas B, Mulley B, Fulton I, Hoffman H. Virtual reality as a pediatric pain modulation technique: a case study. Cyberpsychol Behav 2003 Dec;6(6):633-638. [doi: 10.1089/109493103322725405] [Medline: 14756928]
59. Rao D, Havale R, Nagaraj M, Karobari N, Latha A, Tharay N. Assessment of Efficacy of Virtual Reality Distraction in Reducing Pain Perception and Anxiety in Children Aged 6-10 Years: A Behavioral Interventional Study. International Journal of Clinical Pediatric Dentistry 2019;12(6):512-513. [doi: 10.5005/jp-journals-10005-1694] [Medline: 32440065]
60. Giménez N, Pedrazas D, Redondo S, Quintana S. [Informed consent process in clinical trials: Insights of researchers, patients and general practitioners]. Aten Primaria 2016 Oct;48(8):518-526 [FREE Full text] [doi: 10.1016/j.aprim.2015.10.008] [Medline: 26777978]
61. McLeod C, Norman R, Litton E, Saville BR, Webb S, Snelling TL. Choosing primary endpoints for clinical trials of health care interventions. Contemp Clin Trials Commun 2019 Dec;16:100486 [FREE Full text] [doi: 10.1016/j.conctc.2019.100486] [Medline: 31799474]
62. von Niederhäusern B, Magnin A, Pauli-Magnus C. The impact of clinical trial units on the value of clinical research in Switzerland. Swiss Med Wkly 2018;148:w14615 [FREE Full text] [doi: 10.4414/smw.2018.14615] [Medline: 29698540]
63. Barclay S, Moran E, Boase S, Johnson M, Lovick R, Graffy J, et al. Primary palliative care research: opportunities and challenges. BMJ Support Palliat Care 2019 Dec;9(4):468-472 [FREE Full text] [doi: 10.1136/bmjspcare-2018-001653] [Medline: 30755396]
64. Devouge E, Biard M, Beuzeboc J, Tavolacci M, Schuers M. Motivations and willingness of general practitioners in France to participate in primary care research as investigators. Fam Pract 2019 Oct;36(5):552-559. [doi: 10.1093/fampra/cmy126] [Medline: 30605509]
65. Díez Domingo J. Investigación en pediatría de atención primaria: la asignatura pendiente. Anales de Pediatría 2004;61(4):289-291. [doi: 10.1016/s1695-4033(04)78390-3]
66. Giménez N, Caro C, Ponsa E, Perez Ortiz AM, Navazo I, Gavagnach M. Rising to the Challenge of promoting research in primary care and nursing: Research productivity and professional view. Enfermería Clínica (English Edition) 2017 May;27(3):144-152. [doi: 10.1016/j.enfcli.2017.03.008] [Medline: 28434929]
67. Violán Fors C, GrandesOdriozola G, Zabaleta-del-Olmo E, Gavilán Moral E. [Research in primary care as an area of knowledge. SESPAS Report 2012]. Gac Sanit 2012 Mar;26 Suppl 1:76-81 [FREE Full text] [doi: 10.1016/j.gaceta.2011.05.018] [Medline: 22047623]
68. Ellerton K, Tharmarajah H, Medres R, Brown L, Ringelblum D, Vogel K, et al. The VRIMM study: Virtual Reality for IMMunisation pain in young children-protocol for a randomised controlled trial. BMJ Open 2020 Aug 20;10(8):e038354 [FREE Full text] [doi: 10.1136/bmjopen-2020-038354] [Medline: 32819997]
69. García-Garcés L, Ruiz-Zaldibar C, Lluesma-Vidal M, Serra-Guillén I, Tomás-Saura I, Sánchez-López M. The effect of virtual reality versus standard-of-care treatment on pain perception during paediatric vaccination: study protocol for a randomised controlled trial. J Clin Nurs (forthcoming) 2022.
70. Uman LS, Chambers CT, McGrath PJ, Kisely S. A systematic review of randomized controlled trials examining psychological interventions for needle-related procedural pain and distress in children and adolescents: an abbreviated cochrane review. J Pediatr Psychol 2008 Sep;33(8):842-854 [FREE Full text] [doi: 10.1093/jpepsy/jsn031] [Medline: 18387963]
71. Taddio A, Ipp M, Thivakaran S, Jamal A, Parikh C, Smart S, et al. Survey of the prevalence of immunization non-compliance due to needle fears in children and adults. Vaccine 2012 Jul 06;30(32):4807-4812. [doi: 10.1016/j.vaccine.2012.05.011] [Medline: 22617633]
72. Manne S, Jacobsen P, Redd W. Assessment of acute pediatric pain: do child self-report, parent ratings, and nurse ratings measure the same phenomenon? Pain 1992 Jan;48(1):45-52. [doi: 10.1016/0304-3959(92)90130-4] [Medline: 1738574]
73. Simons LE, Sieberg CB, Carpino E, Logan D, Berde C. The Fear of Pain Questionnaire (FOPQ): assessment of pain-related fear among children and adolescents with chronic pain. J Pain 2011 Jun;12(6):677-686. [doi: 10.1016/j.jpain.2010.12.008] [Medline: 21354866]
74. Dahlquist LM, Busby SM, Slifer KJ, Tucker CL, Eischen S, Hilley L, et al. Distraction for children of different ages who undergo repeated needle sticks. J Pediatr Oncol Nurs 2002;19(1):22-34. [doi: 10.1053/jpon.2002.30009] [Medline: 11813138]

XSL•FO
RenderX

75. Stewart JL. "Getting used to it": children finding the ordinary and routine in the uncertain context of cancer. Qual Health Res 2003 Mar;13(3):394-407. [doi: 10.1177/1049732302250336] [Medline: 12669339]
76. Uman LS, Chambers CT, McGrath PJ, Kisely S, Matthews D, Hayton K. Assessing the quality of randomized controlled trials examining psychological interventions for pediatric procedural pain: recommendations for quality improvement. J Pediatr Psychol 2010 Aug;35(7):693-703. [doi: 10.1093/jpepsy/jsp104] [Medline: 19966314]
77. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013 May;14(5):365-376. [doi: 10.1038/nrn3475] [Medline: 23571845]
78. Uman L, Birnie K, Noel M, Parker J, Chambers C, Mcgrath P. Psychological interventions for needle-related procedural pain and distress in children and adolescents. Cochrane Database of Systematic Reviews 2013 Oct:CD005179. [doi: 10.1002/14651858.CD005179.pub3] [Medline: 24108531]
79. Hopewell S, Loudon K, Clarke M, Oxman A, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009 Jan:MR000006. [doi: 10.1002/14651858.MR000006.pub3] [Medline: 19160345]

Abbreviations

CASPe: Critical Appraisal Skills Program guide in Spanish
IV: inverse variance
RCT: randomized controlled trial
VAS: visual analog scale
VR: virtual reality
WBFPS: Wong-Baker Faces Pain Rating Scale

Edited by N Zary; submitted 19.11.21; peer-reviewed by B Nievas Soriano, KA Nguyen; comments to author 22.01.22; revised version received 19.03.22; accepted 13.05.22; published 09.08.22.

Please cite as:
Lluesma-Vidal M, Carcelén González R, García-Garcés L, Sánchez-López MI, Peyro L, Ruiz-Zaldibar C
Effect of Virtual Reality on Pediatric Pain and Fear During Procedures Involving Needles: Systematic Review and Meta-analysis
JMIR Serious Games 2022;10(3):e35008
URL: https://games.jmir.org/2022/3/e35008
doi:10.2196/35008
PMID:35943776

©Marta Lluesma-Vidal, Raquel Carcelén González, Laura García-Garcés, María I Sánchez-López, Loreto Peyro, Cayetana Ruiz-Zaldibar. Originally published in JMIR Serious Games (https://games.jmir.org), 09.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Effectiveness of Serious Games in Improving Memory Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis

Alaa Abd-alrazaq[1], PhD; Dari Alhuwail[2,3], PhD; Eiman Al-Jafar[4], PhD; Arfan Ahmed[1], PhD; Farag Shuweihdi[5], PhD; Shuja Mohd Reagu[6], PhD; Mowafa Househ[7], PhD

[1]AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

[2]Information Science Department, Kuwait University, Kuwait, Kuwait

[3]Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait

[4]Kuwait Health Informatics Association, Kuwait, Kuwait

[5]Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

[6]Mental Health Services, Hamad Medical Corporation, Doha, Qatar

[7]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University, Qatar Foundation
P.O. Box 34110, Doha Al Luqta St, Ar-Rayyan
Doha
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Memory, one of the main cognitive functions, is known to decline with age. Serious games have been used for improving memory in older adults. The effectiveness of serious games in improving memory has been assessed by many studies. To draw definitive conclusions about the effectiveness of serious games, the findings of these studies need to be pooled and aggregated.

Objective: This study aimed to assess the effectiveness of serious games in improving memory in older adults with cognitive impairment.

Methods: A systematic review of randomized controlled trials was carried out. The search sources included 8 databases, the reference lists of the included studies and relevant reviews, and the studies that cited the included studies. In total, 2 reviewers (AA and MH) independently carried out the study selection, data extraction, risk of bias assessment, and quality of evidence appraisal. Extracted data were synthesized using a narrative approach and a statistical approach (ie, multilevel meta-analysis), as appropriate.

Results: Of the 618 citations retrieved, 18 (2.9%) met the eligibility criteria for this review. Of these 18 studies, 15 (83%) randomized controlled trials were included in 10 multilevel meta-analyses. We found that serious games were more effective than no or passive interventions in improving nonverbal memory (*P*=.02; standardized mean difference [SMD]=0.46, 95% CI 0.09-0.83) and working memory (*P*=.04; SMD=0.31, 95% CI 0.01-0.60) but not verbal memory (*P*=.13; SMD=0.39, 95% CI −0.11 to 0.89). The review also showed that serious games were more effective than conventional exercises in improving verbal memory (*P*=.003; SMD=0.46, 95% CI 0.16-0.77) but not nonverbal memory (*P*=.30; SMD=−0.19, 95% CI −0.54 to 0.17) or working memory (*P*=.99; SMD=0.00, 95% CI −0.45 to 0.45). Serious games were as effective as conventional cognitive activities in improving verbal memory (*P*=.14; SMD=0.66, 95% CI −0.21 to 1.54), nonverbal memory (*P*=.94; SMD=−0.01, 95% CI −0.32 to 0.30), and working memory (*P*=.08; SMD=0.37, 95% CI −0.05 to 0.78) among older adults with cognitive impairment. Finally, the effect of adaptive serious games on working memory was comparable with that of nonadaptive serious games (*P*=.08; SMD=0.18, 95% CI −0.02 to 0.37).

XSL•FO
RenderX

Conclusions: Serious games have the potential to improve verbal, nonverbal, and working memory in older adults with cognitive impairment. However, our findings should be interpreted cautiously given that most meta-analyses were based on a few studies (≤3) and judged to have a low quality of evidence. Therefore, serious games should be offered as a supplement to existing proven and safe interventions rather than as a complete substitute until further, more robust evidence is available. Future studies should investigate the short- and long-term effects of serious games on memory and other cognitive abilities among people of different age groups with or without cognitive impairment.

(JMIR Serious Games 2022;10(3):e35202) doi:10.2196/35202

KEYWORDS

serious games; cognitive training; exergames; mild cognitive impairment; Alzheimer disease; dementia; memory; systematic reviews; meta-analysis; mobile phone

Introduction

Background

Life expectancy has increased worldwide as people have better access to health care services and an improved standard of living. As a result, people are living longer [1-3]. According to the United Nations World Population Aging 2020 report [4], the number of people aged ≥65 years has increased up to 727 million worldwide. The older population group is expected to increase to 16% by 2050 compared with 9.3% in 2020 [4]. The older population group is more likely to develop cognitive impairment [5,6], which is a decline in cognitive abilities and functions such as memory, attention, concentration, learning, and language [7,8]. According to the Alzheimer's Association, approximately 12% to 18% of people aged ≥60 years have mild cognitive impairment (MCI) [9].

MCI refers to a decline in the ability to learn new information or recall stored information and occurs along a continuum that ranges from normal to severely impaired cognition [10]. Although inconsistencies exist in screening for MCIs, it is certain that they occur because of brain changes owing to multiple factors, including older age, injuries to the brain, diabetes, hypertension, stroke, depression, and physical inactivity [11]. Memory is one of the main cognitive functions that decline with age. Memory is known as the ability of the brain to hold information and recall it as needed. There are different types of memory: verbal, nonverbal, and working memory. Verbal memory refers to a person's ability to remember what they read or hear of information that was already learned [12]. On the other hand, nonverbal memory refers to storing, retrieving, and remembering nonverbal information, content, or experiences, such as images, feelings, tastes, sounds, shapes, and smells [13]. Furthermore, memory is divided into 3 types according to the period for which the memorized information is retained: short-term, long-term, and working memory. Short-term memory temporarily holds a limited amount of information [14], whereas long-term memory refers to the relatively permanent storage and recall of information [15]. Working memory refers to the temporary storage of a limited amount of information to be used in the execution of cognitive activities such as learning, reasoning, and comprehension [16].

Several nonpharmacological interventions can be used to improve memory, such as physical exercise, cognitive behavioral therapy, psychosocial therapy, good nutrition, and serious games [17]. Serious games are defined as electronic games that are played for purposes beyond leisure to promote the users' mental, physical, and social well-being [18,19]. Recent evidence suggests that exergames are effective in improving physical and cognitive function in people with MCIs [20] as well as their compliance and adherence to medical interventions embedded in serious games [21,22]. Previous systematic reviews have shown that serious games have the potential to prevent or alleviate mental disorders such as depression [23], anxiety [24], and cognitive impairment [25]. Several types of serious games have been used to improve cognitive abilities, namely (1) cognitive training games (which deliver cognitive activities to maintain or improve cognitive functions) and (2) exergames (which entail physical exercises as part of the intended gameplay [25]). Compared with conventional exercise and cognitive training, serious games can positively affect mood, social functioning, mental health well-being, and cognitive flexibility in older adults [26-29].

Research Problem and Objectives

The effectiveness of serious games in improving memory has been assessed by many studies. To draw definitive conclusions about the effectiveness of serious games, the findings of these studies need to be pooled and aggregated. Several systematic reviews have summarized the evidence from these studies; however, they had a different aim and scope from this review. Specifically, these reviews (1) focused on healthy older adults and not necessarily those with cognitive impairment [17,30-33] (therefore, future reviews should consider older adults with cognitive impairment), (2) included pilot randomized controlled trials (RCTs) and quasi-experiments [17,20,33,34] (thus, future reviews should include only RCTs), (3) performed an outdated search (>5 years [17,32,34]; therefore, an updated review or a new review are required), (4) did not assess the quality of evidence [17,20,30,33,34] (thus, the quality of the evidence should be assessed in future reviews), (5) only focused on a specific type of serious game such as cognitive training games [30,34] and exergames [17,20,33] (hence, future reviews should consider all types of serious games), (6) focused on a certain type of memory (working memory [34]; therefore, all types of memory should be considered in upcoming reviews), or (7) did not compare the effect of serious games with a specific comparator (eg, no intervention, conventional exercises, or conventional cognitive activities [17,20,30,33,34]; thus, further reviews are needed to compare the effect of serious games with a specific comparator). To address the aforementioned gaps, this study aimed to assess the effectiveness of serious games in

XSL•FO
RenderX

improving memory among older adults with cognitive impairment. This review focused only on memory as other cognitive domains—for example, global cognition [25], executive functions [35], and processing speed [36]—were targeted by previous reviews.

Methods

The authors followed the expanded version of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to conduct a systematic review and meta-analyses (Multimedia Appendix 1). The protocol for this review was registered with PROSPERO (CRD42021292150).

Eligibility Criteria

This review included only RCTs that looked at the effectiveness of serious games in improving memory in older adults with cognitive impairment. The target intervention in this review was serious games supplied on any digital platform, such as computers (PCs), consoles (Xbox and PlayStation), mobile phones, handheld devices, Nintendo, or any other computerized device. Furthermore, components of gaming had to be used as an important and major technique for reaching the intervention's goal. Serious games had to be used solely for the purpose of therapy. Studies combining serious games with other interventions were eligible if the control group received the same adjacent intervention. Nondigital games and those used for other purposes, such as monitoring, screening, diagnosis, and research, were excluded.

The study focused on the older adult population (aged ≥60 years) who had any type of cognitive impairment or condition (eg, MCI, Alzheimer disease, or dementia). Their diagnosis had to be confirmed by checking the inclusion criteria or baseline scores against standardized diagnostic criteria (eg, Mini-Mental State Examination and Montreal Cognitive Assessment). This review did not focus on healthy older adults, health care providers, or caregivers. No restrictions were applied regarding sex and ethnicity.

The main outcome of interest in this review was memory regardless of the type (verbal, nonverbal, or working memory) and regardless of the tool used for measuring the outcome. Studies were excluded if they assessed only other cognitive outcomes (eg, language and processing speed), cost-effectiveness, acceptance, feasibility, or satisfaction. This review focused on outcome data that were measured immediately after the intervention rather than on follow-up data.

Only RCTs conducted in English and from 2010 onward were considered. Pilot or feasibility RCTs, quasi-experiments, observational studies, and reviews were omitted. Studies published as journal articles, conference proceedings, or dissertations were included. Reviews, conference abstracts, proposals, editorials, and commentaries were all excluded. Finally, no restrictions related to the country of publication, comparator, or study setting were applied.

Information Sources and Search Strategy

The studies that were relevant to this review were found by searching 7 bibliographic databases: MEDLINE (via Ovid), PsycINFO (via Ovid), EMBASE (via Ovid), CINAHL (via EBSCO), IEEE Xplore, ACM Digital Library, and Scopus. Furthermore, we searched the search engine Google Scholar. Owing to the high number of papers obtained through Google Scholar, only the first 10 pages (ie, 100 records) were taken into account as they were automatically ordered based on their relevance [37]. The first author (AA) conducted the search on August 6, 2021. An automatic alert was set up to retrieve studies that were added to the databases after that date; this continued for 16 weeks (ending on December 5, 2021). Forward reference list checking (ie, screening studies that cited the included studies) and backward reference list checking (ie, screening the reference lists of the included studies and relevant reviews) were carried out to retrieve further studies.

To develop the search query for this review, the authors consulted 2 experts in digital mental health and checked the search queries used in other systematic reviews within this field. The chosen search terms were related to the target population (eg, cognitive impairment), target intervention (eg, serious games and exergames), and target study design (eg, RCTs). Multimedia Appendix 2 summarizes the search query that was used for searching each of the 8 databases.

Selection Process

Relevant studies were identified taking the following steps. First, the obtained studies were imported into EndNote X8 (Clarivate Analytics) to identify and delete duplicate items. Second, the titles and abstracts of all the retrieved studies were evaluated in the second phase by 2 reviewers (AA and MH) working independently. Finally, the 2 reviewers independently evaluated the entire texts of the studies included in the previous step. Any disagreements in the 2 previous steps were resolved via discussion. The interrater agreement (Cohen κ) in steps 2 and 3 was 0.94 and 0.96, respectively, indicating a near-perfect level of interrater agreement [38].

Data Collection Process

In total, 2 independent reviewers (AA and MH) used Microsoft Excel to extract data from all the included studies. The data extraction form used to extract data from the included studies was pilot-tested using 2 of the included studies (Multimedia Appendix 3). The reviewers' disagreements were resolved through discussion. An interrater agreement of 0.85 was observed, indicating a near-perfect degree of agreement. If data such as the mean, SD, and sample size were unavailable from the published studies, contact was made with the first and corresponding authors in an attempt to retrieve them.

Study Risk of Bias Assessment

The Cochrane Collaboration recommends assessing the risk of bias via 2 independent reviewers (AA and MH) using the Risk of Bias 2 (RoB 2) tool [39]; as such, these guidelines were followed for this review. The RoB 2 tool assesses the risk of bias in 5 domains of RCTs: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result [39]. The risk of bias judgments in these domains were used to determine the overall risk of bias of each included study. Any inconsistencies in decisions between the reviewers were resolved

XSL•FO
RenderX

by consulting a third reviewer. Interrater agreement between the reviewers was near perfect (Cohen κ=0.93) [38].

Synthesis Methods

A narrative and statistical approach was used to synthesize the information acquired. In our narrative synthesis, we used texts and tables to describe the characteristics of the included studies (demographic, intervention, comparator, and outcome variables). The results of the experiments were categorized and pooled based on measured outcome (ie, verbal, nonverbal, and working memory) and the comparator (ie, control, conventional exercises, conventional cognitive training, and other serious games). A meta-analysis was conducted when at least two studies with the same measured outcome and comparator submitted enough data (ie, mean, SD, and number of participants in each intervention group). Owing to the type of data for the outcome of interest (memory) being continuous and the methods used to measure the outcome being variable throughout the included studies, the standardized mean difference (SMD; Cohen *d*) was used to analyze the overall effect of each study. The random effects model was used for the analysis because of the high clinical heterogeneity among the meta-analyzed studies in terms of serious game characteristics (eg, type, duration, frequency, and period), population characteristics (eg, sample size, mean age, and health condition), and outcome measures (ie, tools and follow-up period). As several studies used more than one outcome measure to assess memory, the dependency on effect sizes within or across studies will be introduced in the meta-analysis. As a result, a multilevel meta-analysis considering the dependency on effect sizes and sampling covariance between the effect sizes was used [40-42]. Namely, the multilevel meta-analysis should be applied when effect sizes within the same study are very likely to be more similar to each other than the effect sizes across studies [42]. The *R* (version 4.3.1; R Foundation for Statistical Computing) statistical package was used to perform the analysis. We used the function *rma.mv* in the library *metafor*, which is a library in *R*, to perform the multilevel meta-analysis [43].

If we observed a statistically significant difference between the groups in a meta-analysis, we further sought to examine if it was clinically significant. The phrase "minimal clinically important difference" (MCID) refers to the smallest change in a measured outcome that a patient would consider worthwhile and significant enough to warrant a change in treatment. The MCID boundaries were calculated as 0.5 times the SMD of the meta-analyzed studies.

We calculated I^2 and a chi-square *P* value to investigate the degree and statistical significance of the heterogeneity in the meta-analyzed studies, respectively. A chi-square *P* value of ≤.05 suggests heterogeneous meta-analyzed studies [44]. When I^2 ranged from 0% to 40%, 30% to 60%, 50% to 90%, and 75% to 100%, the degree of heterogeneity was judged to be insignificant, moderate, substantial, or considerable, respectively [44].

Certainty of Evidence

To appraise the overall quality of evidence resulting from the meta-analyses, we applied the Grading of Recommendations Assessment, Development, and Evaluation approach [45], which assesses the quality of evidence based on 5 domains: risk of bias, inconsistency (ie, heterogeneity), indirectness, imprecision, and publication bias [45]. In total, 2 reviewers independently rated the overall quality of the meta-analyzed evidence, and disagreements were resolved through discussion. The interrater agreement of the reviewers was considered near perfect (Cohen κ=0.87) [38].

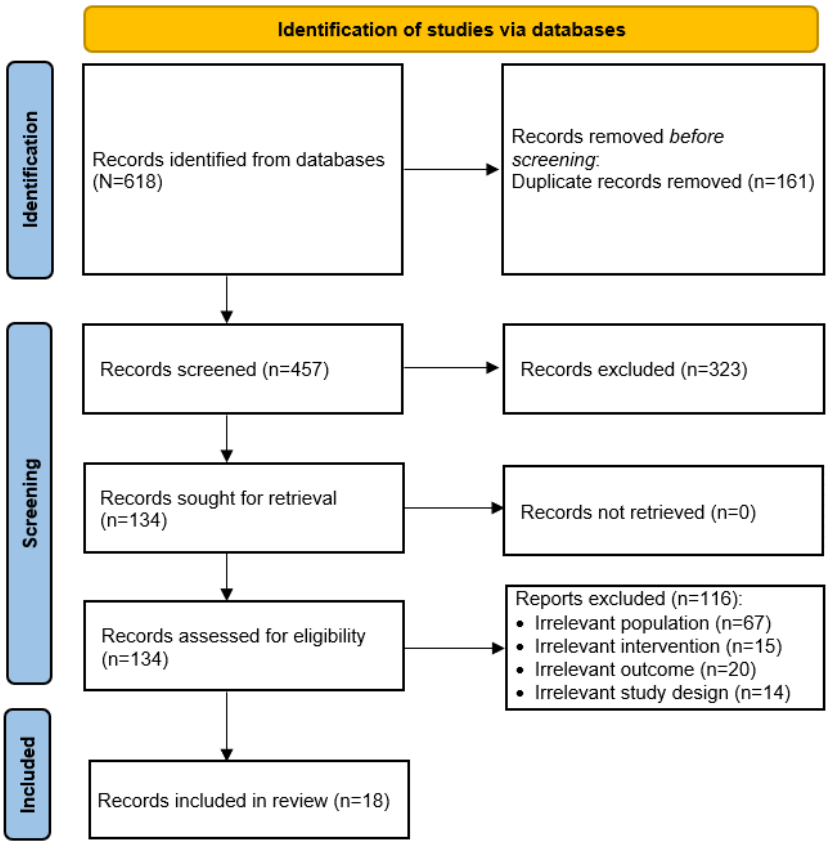
Results

Study Selection

The total number of records retrieved by searching the predefined databases was 618 (Figure 1). Of these 618 records, 161 (26.1%) duplicates were removed using the EndNote software. Checking titles and abstracts of the remaining records led to the exclusion of 52.3% (323/618). After reading the full texts of the remaining 134 publications, 116 (86.6%) were excluded, mainly because of the population (n=67, 57.8%). The list of studies that were excluded after screening the full texts is provided in Multimedia Appendix 4. No additional studies were found through backward and forward reference list checking. In total, 18 RCTs were included in this review [46-63]. Of these 18 studies, 15 (83%) were included in 10 meta-analyses [47-49,51,52,54-63]. A total of 17% (3/18) of the studies were excluded from the meta-analyses because 33% (1/3) [46] did not report the data required for the meta-analysis (eg, mean and SD) and 67% (2/3) [53,61] compared serious games with other serious games that had different characteristics; therefore, including them in a meta-analysis would not make sense.

XSL•FO
RenderX

Figure 1. Flowchart of the study selection process.



Study Characteristics

The included studies were published between 2012 and 2021 (Table 1). The year in which the largest number of included studies was published was 2015 (4/18, 22%). The included studies were carried out in 13 different countries, and there was a general equal distribution of studies in these countries. All the included studies were peer-reviewed journal articles except for a book chapter included (1/18, 6%). The trial type was parallel RCT in most of the included studies (17/18, 94%).

The sample size of the included studies varied from 20 to 209, with an average of 81. The mean age of the participants in the included studies ranged from 66 to 83.1 years, with an average of 74.5 years. The percentage of men in the included studies ranged from 21.5% to 71%, with an average of 46.5%. The participants in most of the included studies had MCI (14/18, 78%). Participants were recruited from clinical settings in 67% (12/18) of the studies, from the community in 28% (5/18) of the studies, and from both clinical settings and the community in 6% (1/18) of the studies.

Serious games alone were used as interventions in 89% (16/18) of the included studies, whereas the remaining 11% (2/18) of the studies used serious games combined with conventional exercises [48] or sham exercises [49] (Table 2). The included studies used 16 different serious games. On the basis of the therapeutic modality that they delivered, the serious games used in the included studies were grouped into 2 types: cognitive training games (16/18, 89%) and exergames (2/18, 11%). Games were designed with a “serious” purpose from the beginning (designed serious games) in all studies except for 6% (1/18) that used a purpose-shifted game (which was not designed as a serious game from the start but rather was used for a serious purpose). The most common platform used for playing the games were computers (14/18, 78%). In 67% (12/18) of the studies, serious games were played under the supervision of health care providers or caregivers. The duration of the games in the included studies ranged from 7 to 90 minutes, and the most common duration was 60 minutes (7/18, 39%). The frequency of playing the games varied between 2 and 7 times per week, but it was 2 times per week in half of the studies (9/18, 50%). The period of intervention ranged from 2 to 25 weeks, but it was ≤12 weeks in 72% (13/18) of the studies.

The comparison groups received only passive interventions in 39% (7/18) of the studies, whereas they received only active interventions in 44% (8/18) of the studies (eg, conventional exercises and conventional cognitive activities; Table 3). In total, 17% (3/18) of the studies delivered both active and passive interventions as comparators. The duration of the active comparators ranged from 7 to 100 minutes. The frequency of the active comparators varied between 2 and 7 times per week. The period of the active comparators varied between 2 and 25 weeks. Most of the included studies (16/18, 89%) measured more than one outcome. The measured outcomes were verbal memory in 78% (14/18) of the studies, nonverbal memory in 61% (11/18) of the studies, and working memory in 67% (12/18) of the studies. The studies used 32 different tools to measure these outcomes, but the most common tool used was the Wechsler Memory Scale Third Edition (7/18, 39%). The outcomes were measured immediately after the intervention in all the included studies (18/18, 100%). The follow-up period

XSL•FO
RenderX

ranged from 4 to 264 weeks. Participant attrition was reported in 89% (16/18) of the studies, and it ranged from 0 to 23.

Table 1. Characteristics of the studies and populations (N=18).

Study	Year	Country	Publication type	RCT[a] type	Sample size	Age, mean	Sex (male; %)	Health condition	Setting
Valdes et al [46]	2012	United States	Journal article	Parallel	195	77.7	33.3	MCI[b]	Clinical
Zhuang et al [47]	2013	China	Journal article	Parallel	33	83.1	24.2	MCI; dementia	Clinical
Hagovská et al [48]	2016	Slovakia	Journal article	Parallel	80	67	51.2	MCI	Clinical
Singh et al [49]	2014	Australia	Journal article	Factorial	100	70.1	32	MCI	Community
Gooding et al [50]	2016	United States	Journal article	Parallel	96	75.6	58.1	MCI	Clinical
Liao et al [51]	2021	Taiwan	Journal article	Parallel	61	81.5	32.6	MCI	Community
Finn and McDonald [52]	2015	Australia	Journal article	Parallel	31	75.6	71	MCI	Clinical
Park and Park [53]	2017	South Korea	Journal article	Parallel	78	67.3	53.8	MCI	Community
Cavallo et al [54]	2016	Italy	Journal article	Parallel	80	76.4	36.3	AD[c]	Clinical
Leung et al [55]	2015	Hong Kong	Journal article	Parallel	209	70.1	21.5	MCI	Community
Yang and Kwak [56]	2017	South Korea	Journal article	Parallel	20	71	70	AD	Clinical
Tarnanas et al [57]	2014	Greece	Book chapter	Parallel	114	70.3	39	MCI	Clinical
Flak et al [58]	2019	Norway	Journal article	Parallel	85	66	66.7	MCI	Clinical
Herrera et al [59]	2012	France	Journal article	Parallel	22	76.6	50	MCI	Clinical
Savulich et al [60]	2017	United Kingdom	Journal article	Parallel	42	76.1	59.5	MCI	Clinical
Boller et al [61]	2012	France	Journal article	Parallel	36	81.2	36.1	AD	Clinical
Karssemeijer et al [62]	2019	Netherlands	Journal article	Parallel	115	79.9	53.9	Dementia	Clinical, community
Hyer et al [63]	2015	United States	Journal article	Parallel	68	75.2	47.1	MCI	Community

[a]RCT: randomized controlled trial.

[b]MCI: mild cognitive impairment.

[c]AD: Alzheimer disease.

XSL•FO
RenderX

Table 2. Characteristics of the interventions (N=18).

Study	Serious game name	Serious game type	Platform	Supervision	Duration (minutes)	Frequency (times per week)	Period (weeks)
Valdes et al [46]	SOPT	Cognitive training game	PC	Supervised	60	2	5
Zhuang et al [47]	NR[a]	Cognitive training game	PC	Supervised	75	3	24
Hagovská et al [48]	CogniPlus	Cognitive training game	PC	Supervised and unsupervised	30	2	10
Singh et al [49]	COGPACK	Cognitive training game	PC	Supervised	75	2	25
Gooding et al [50]	BrainFitness	Cognitive training game	PC	Supervised and unsupervised	60	2	17
Liao et al [51]	Tano and Long-Good	Exergame	Kinect, VR[b] headset	Supervised	60	3	12
Finn and McDonald [52]	E-Prime	Cognitive training game	PC	Supervised	NR	2	4
Park and Park [53]	CoTras	Cognitive training game	PC	Supervised	30	3	10
Cavallo et al [54]	Brainer	Cognitive training game	PC	Supervised	30	3	12
Leung et al [55]	BrainFitness	Cognitive training game	PC	Unsupervised	60	3	13
Yang and Kwak [56]	Brain-Care	Cognitive training game	PC	Unsupervised	60	2	12
Tarnanas et al [57]	Virtual Reality Museum	Cognitive training game	VR headset	Supervised	90	2	21
Flak et al [58]	Cogmed	Cognitive training game	PC	Unsupervised	30-40	5	5
Herrera et al [59]	NR	Cognitive training game	PC	Supervised	60	2	12
Savulich et al [60]	Game Show	Cognitive training game	Tablet	Supervised	60	2	4
Boller et al [61]	NR	Cognitive training game	PC	Supervised	7-10	3	2
Karssemeijer et al [62]	NR	Exergame	Stationary bicycle and screen	Supervised	30-50	3	12
Hyer et al [63]	Cogmed	Cognitive training game	PC	Supervised and unsupervised	40	7	5-7

[a]NR: not reported.

[b]VR: virtual reality.

XSL•FO
RenderX

Table 3. Characteristics of the comparators and outcomes (N=18).

Study	Comparator	Duration (minutes)	Frequency (times per week)	Period (weeks)	Measured outcomes	Outcome measures	Follow-up	Attrition, N
Valdes et al [46]	Control	N/A[a]	N/A	N/A	VM[b]	HVLT[c]; RAVLT[d]; RBMT[e]	After the intervention; 52-, 104-, 156-, and 261-week follow-up	NR[f]
Zhuang et al [47]	Control	N/A	N/A	N/A	VM	ACE-R[g]	After the intervention	10
Hagovská et al [48]	Conventional exercises	30	7	10	VM	ACE-R	After the intervention	2
Singh et al [49]	Conventional exercises+sham cognitive training; serious games+conventional exercises; control	Conventional exercises+sham cognitive training: 75; serious games+conventional exercises: 100; control: 60	2	25	VM; NVM[h]	BVRT-R[i]; WMS-III-LM[j]	After the intervention; 74-week follow-up	14
Gooding et al [50]	Empirically validated serious game; commercially available serious game	60	2	17	VM; NVM	WMS-R-VR-II[k]; WMS-R-LM[l]; BSRT[m]	After the intervention	22
Liao et al [51]	Conventional exercises	60	3	12	VM; WM[n]	CVLT[o]; SBTT[p]	After the intervention	15
Finn and McDonald [52]	Control	NR	2	4	VM; WM	WMS-IV-VPA-II[q]; WMS-IV-SS[r]	After the intervention	7
Park and Park [53]	Commercially available exergame	30	3	10	VM; WM	RAVLT; WAIS-DSB[s]	After the intervention	0
Cavallo et al [54]	Control	N/A	N/A	N/A	VM; NVM; WM	RBMT; WMS-R-DSB[t]; TSWRT[u]	After the intervention; 24-week follow-up	4
Leung et al [55]	Control	60	3	13	VM; NVM; WM	WMS-III-FP[v]; WMS-III-LM; WMS-III-DST[w]; WMS-III-VSST[x]	After the intervention	0
Yang and Kwak [56]	Control	N/A	N/A	N/A	VM; NVM; WM	ROCFT[y]; SVLT[z]; WMS-III-DSB[aa]	After the intervention	0
Tarnanas et al [57]	Control; conventional cognitive activities	90	2	21	VM; NVM; WM	ROCFT; RAVLT; WMS-III-DSB	After the intervention	9

XSL•FO
RenderX

Study	Comparator	Duration (minutes)	Frequency (times per week)	Period (weeks)	Measured outcomes	Outcome measures	Follow-up	Attrition, N
Flak et al [58]	Nonadaptive serious game	30 to 40	5	5	VM; NVM; WM	ROCFT; WMS-III-FII[ab]; WMS-III-LM; CVLT-II[ac]; WMS-III-DSB; WMS-III-SS[ad]; WMS-III-LNS[ae]	After the intervention; 4- and 16-week follow-up	17
Herrera et al [59]	Conventional cognitive activities	60	2	12	VM; NVM; WM	ROCFT-R[af]; BEM-WLTR[ag]; MMSE-R[ah]; WMS-R-DSB	After the intervention; 24-week follow-up	NR
Savulich et al [60]	Control	N/A	N/A	N/A	NVM	BVRT-R	After the intervention	0
Boller et al [61]	Serious game; control	7 to 10	3	2	NVM; WM	SRT[ai]; n-BT[aj]; RST[ak]	After the intervention	0
Karssemeijer et al [62]	Conventional exercises (aerobic exercises); conventional exercises (relaxation and flexibility exercises)	30 to 50	3	12	NVM; WM	LLT-R[al]; WAIS-III-DS[am]; WMS-III-VSST	After the intervention; 24-week follow-up	23

XSL•FO
RenderX

Study	Comparator	Duration (minutes)	Frequency (times per week)	Period (weeks)	Measured outcomes	Outcome measures	Follow-up	Attrition, N
Hyer et al [63]	Nonadaptive serious game	40	7	5 to 7	WM	WMS-III-DST; WMS-III-LNS	After the intervention; 12-week follow-up	9

[a]N/A: not applicable.
[b]VM: verbal memory.
[c]HVLT: Hopkins Verbal Learning Test.
[d]RAVLT: Rey Auditory Verbal Learning Test.
[e]RBMT: Rivermead Behavioral Memory Test.
[f]NR: not reported.
[g]ACE-R: Addenbrooke Cognitive Examination-Revised.
[h]NVM: nonverbal memory.
[i]BVRT-R: Benton Visual Retention Test-Revised, Fifth Edition.
[j]WMS-III-LM: Wechsler Memory Scale Third Edition-Logical Memory.
[k]WMS-R-VR-II: Wechsler Memory Scale-Revised-Visual Reproductions II.
[l]WMS-R-LM: Wechsler Memory Scale-Revised-Logical Memory.
[m]BSRT: Buschke Selective Reminding Test.
[n]WM: working memory.
[o]CVLT: California Verbal Learning Test.
[p]SBTT: spatial n-back task test.
[q]WMS-IV-VPA-II: Wechsler Memory Scale Fourth Edition-Verbal Paired Associates II.
[r]WMS-IV-SS: Wechsler Memory Scale Fourth Edition-Symbol Span.
[s]WAIS-DSB: Wechsler Adult Intelligence Scale-Digit Span Backwards.
[t]WMS-R-DSB: Wechsler Memory Scale-Revised-Digit Span Backwards.
[u]TSWRT: two-syllable word repetition test.
[v]WMS-III-FP: Wechsler Memory Scale Third Edition-Family Pictures.
[w]WMS-III-DST: Wechsler Memory Scale Third Edition-Digit Span Test.
[x]WMS-III-VSST: Wechsler Memory Scale Third Edition-Visual-Spatial Span Test.
[y]ROCFT: Rey-Osterrieth complex figure test.
[z]SVLT: Seoul Verbal Learning Test.
[aa]WMS-III-DSB: Wechsler Memory Scale Third Edition-Digit Span Backwards Test.
[ab]WMS-III-FII: Wechsler Memory Scale Third Edition-Faces II.
[ac]CVLT-II: California Verbal Learning Test-Second Edition.
[ad]WMS-III-SS: Wechsler Memory Scale Third Edition-Symbol Span.
[ae]WMS-III-LNS: Wechsler Memory Scale Third Edition-Letter-Number Sequencing.
[af]ROCFT-R: Rey-Osterrieth complex figure test-Revised.
[ag]BEM-WLTR: Batterie d'Efficience Mnesique-word list total recall.
[ah]MMSE-R: Mini-Mental State Examination-Recall.
[ai]SRT: source recognition task.
[aj]n-BT: n-back task.
[ak]RST: reading span task.
[al]LLT-R: Location Learning Test-Revised.
[am]WAIS-III-DS: Wechsler Adult Intelligence Scale Second Edition-Digit Span.

Risk of Bias in the Studies

An appropriate random allocation sequence for the randomization process was used in 44% (8/18) of the studies. Researchers in 39% (7/18) of the studies concealed the allocation sequence until participants were assigned to the interventions. The groups were comparable at baseline in all studies (18/18, 100%). Thus, the risk of bias owing to the randomization process was rated as low in only 33% (6/18) of the studies (Figure 2). Participants and those who delivered the interventions were aware of the assigned interventions during the trial in 67% (12/18) and 83% (15/18) of the studies, respectively. None of the studies reported a deviation from the intended intervention because of experimental contexts; however, 11% (2/18) of the studies provided insufficient information to verify if protocol deviations had occurred. Appropriate analysis methods (eg, intention-to-treat analysis) were used in 89% (16/18) of the studies to estimate the effect of the intervention. According to

XSL•FO
RenderX

these judgments, the risk of bias because of deviations from the intended interventions was low in 78% (14/18) of the studies (Figure 2).

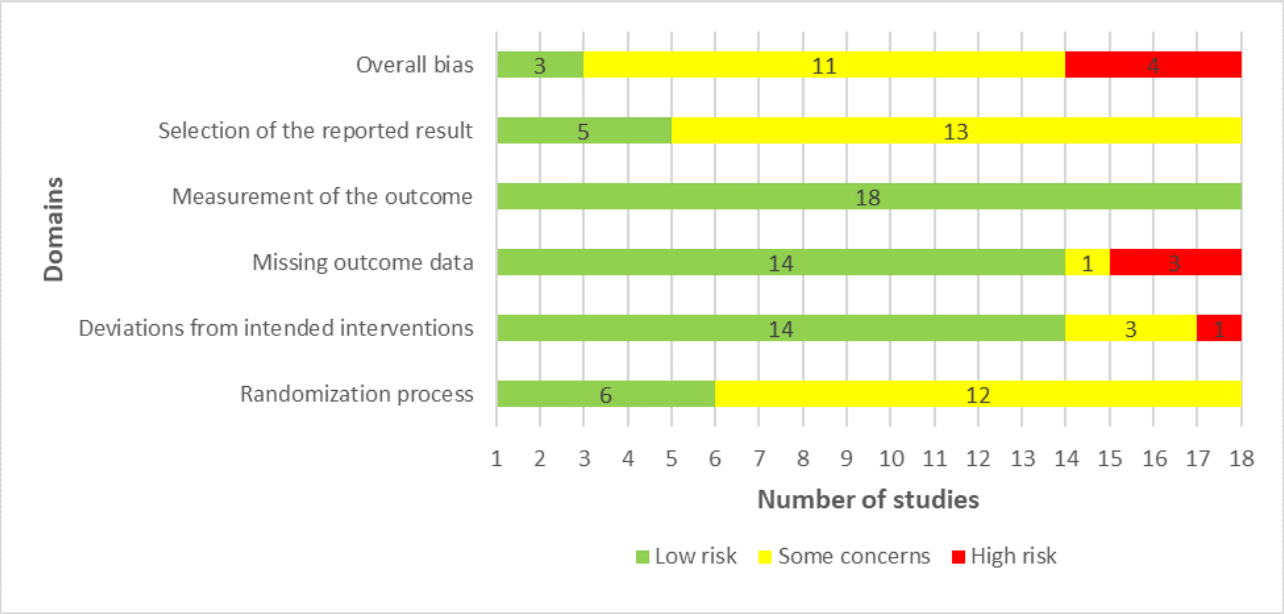
Missing outcome data were <5% in 44% (8/18) of the studies. There was evidence that the findings were not biased by missing outcome data in only 6% (1/18) of the studies. The missing outcome data resulted from reasons that were documented and not related to the outcome in 28% (5/18) of the studies. Therefore, there was a low risk of bias because of missing outcome data in 78% (14/18) of the studies (Figure 2).

In all the included studies (18/18, 100%), the outcomes of interest were evaluated using appropriate measures, and the measurement methods were comparable across the intervention groups. The assessor of the outcome was aware of the assigned interventions in 39% (7/18) of the studies, but it was unlikely that the assessment of the outcome was influenced by knowledge of the intervention received in these studies. Accordingly, all studies (18/18, 100%) had a low risk of bias in the "measuring the outcome" domain (Figure 2).

In total, 28% (5/18) of the studies published their protocols in sufficient detail. In all studies (18/18, 100%), the reported outcome measurements did not differ from those specified in the analysis plan, and there was no evidence that the studies selected their results from many results produced from multiple eligible analyses of the data. On the basis of these judgments, the risk of bias because of the selection of the reported results was considered low in 28% (5/18) of the studies (Figure 2).

In the last domain, "overall bias," the risk of bias was considered high in 22% (4/18) of the studies as they were judged as having a high risk of bias in at least one domain. A total of 61% (11/18) of the studies raised some concerns in the domain of overall bias as they had some issues in at least one of the domains and were not at high risk for any domain. The remaining 17% (3/18) of the studies were judged to be at low risk of bias for the domain of overall bias given that they were rated to be at low risk of bias for all domains. The reviewers' judgments about each "risk of bias" domain for each included study are presented in Multimedia Appendix 5 [46-63].

Figure 2. Review authors' judgments about each "risk of bias" domain.



Results of the Studies

Overview

As mentioned earlier, the included studies assessed the effect of serious games on 3 outcomes: verbal, nonverbal, and working memory. The results of the included studies were divided into 3 groups based on these outcomes. Furthermore, the results for each outcome were grouped based on the comparator used in the studies (ie, control [no or passive interventions], conventional exercises, conventional cognitive activities, and other serious games).

Verbal Memory

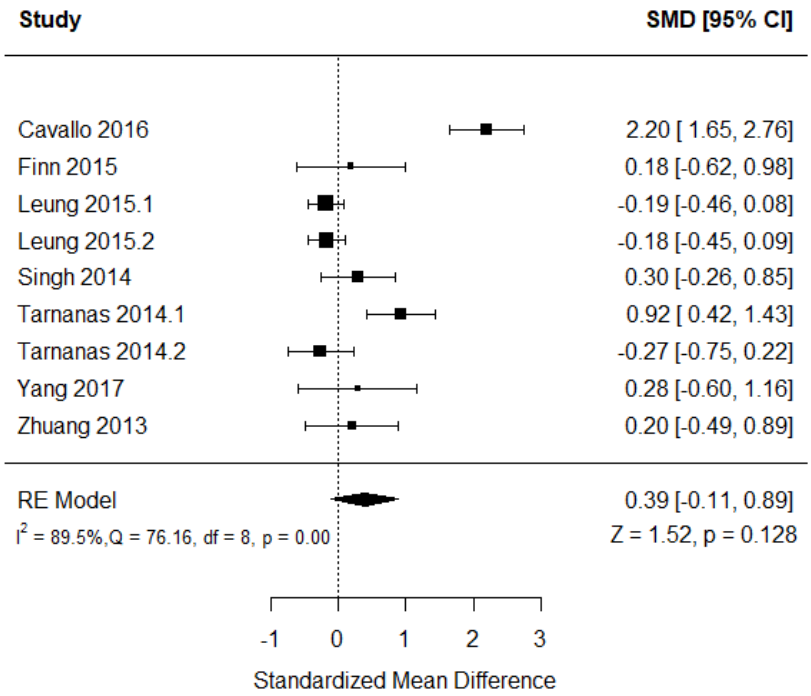
Serious Games Versus Control

The effect of serious games on verbal memory was compared with that of no or passive interventions in 44% (8/18) of the studies [46,47,49,52,54-57]. A total of 13% (1/8) of these studies were not included in the meta-analysis given that they did not report the required data and we could not obtain them when contacting the authors. Of the 7 studies included in the meta-analysis, 2 (29%) assessed verbal memory using 2 different measures [55,57]. Therefore, we included the results of all these measures in the meta-analysis to form 9 comparisons (Figure 3 [47,49,52,54-57]). The meta-analysis showed no statistically significant difference (P=.13) in verbal memory between the serious game and control groups (SMD=0.39, 95% CI −0.11 to 0.89). The statistical heterogeneity of the evidence was considerable (P<.001; I^2=89.5%). The high heterogeneity may be attributed to differences in sample size, participants' health condition, period of the intervention, and outcome measures among the studies included in this analysis. The quality of the evidence was very low as it was downgraded by 5 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 6).

XSL•FO
RenderX

We conducted subgroup analyses, also known as moderator analyses [64], to investigate whether different characteristics of the population (ie, sample size, health condition, and recruitment setting) and intervention (ie, delivery method, duration, frequency, and period) moderated the effect of serious games on verbal memory. As shown in Multimedia Appendix 7, there was no statistically significant difference among all characteristics of the population and intervention except for the health condition of the participants (*P*=.003) and the period of the intervention (*P*=.05).

Figure 3. Forest plot of 7 studies (9 comparisons) comparing the effect of serious games with that of control on verbal memory. RE: random effect; SMD: standardized mean difference [49,51,54,56-59].

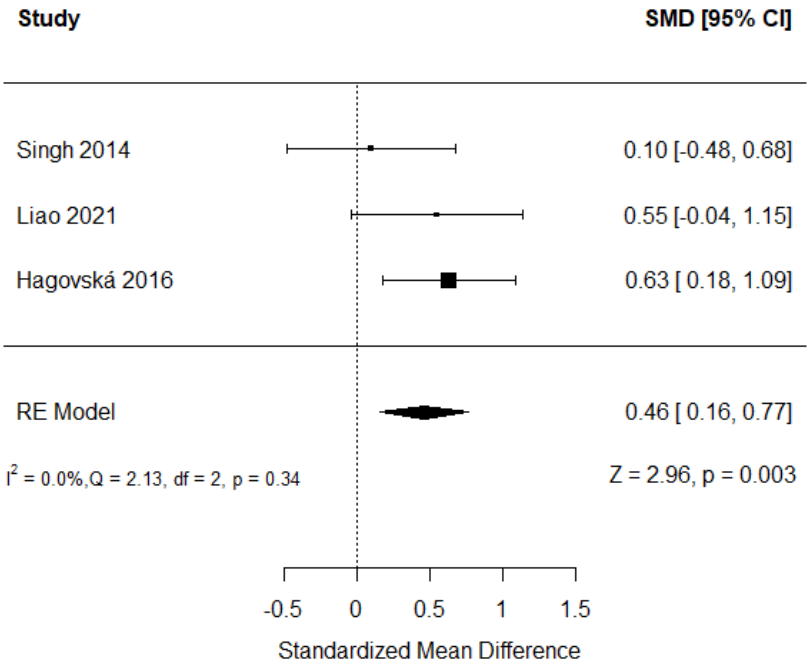


Serious Games Versus Conventional Exercises

The effect of serious games was compared with that of conventional exercises in 17% (3/18) of the studies [48,49,51] (Figure 4 [48,49,51]). A meta-analysis of the results of these studies showed a statistically significant difference in verbal memory (*P*=.003) between the groups, favoring serious games over conventional exercises (SMD=0.46, 95% CI 0.16-0.77). This difference was also clinically important as the overall effect was outside MCID boundaries (−0.23 to 0.23) and its 95% CI did not cross the "no effect" line (zero effect). For this outcome, the MCID boundaries were calculated as –0.5 times to +0.5 times the SMD value (0.46). The statistical heterogeneity of the evidence was not a concern (*P*=.34; I^2=0%). The quality of the evidence was very low as it was downgraded by 3 levels owing to a high risk of bias and imprecision (Multimedia Appendix 6).

XSL•FO
RenderX

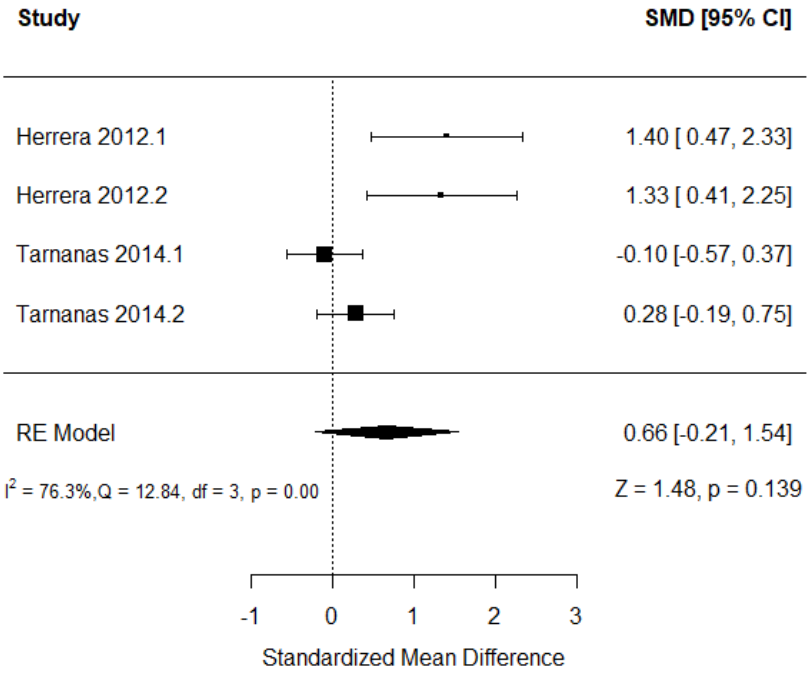
Figure 4. Forest plot of 3 studies comparing the effect of serious games with that of conventional exercises on verbal memory. RE: random effect; SMD: standardized mean difference [50,51,53].



Serious Games Versus Conventional Cognitive Activities

In total, 11% (2/18) of the studies examined the effect of serious games in comparison with conventional cognitive activities [57,59]. These studies assessed verbal memory using 2 different measures. Thus, we included the results of all these measures in a meta-analysis to form 4 comparisons (Figure 5 [57,59]). The meta-analysis showed no statistically significant difference (P=.14) in verbal memory between the groups (SMD=0.66, 95% CI −0.21 to 1.54). The statistical heterogeneity of the evidence was substantial (P<.001; I^2=76.3%). The high heterogeneity may be attributed to differences in the platform of the intervention, period of the intervention, and outcome measures among the studies included in this analysis. The quality of the evidence was very low as it was downgraded by 5 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 6).

Figure 5. Forest plot of 2 studies (4 comparisons) comparing the effect of serious games with that of conventional cognitive activities on verbal memory. RE: random effect; SMD: standardized mean difference [59,61].



XSL•FO
RenderX

Serious Games Versus Other Serious Games

In total, 17% (3/18) of the studies compared the effect of serious games on verbal memory with that of other serious games [50,53,58]. Specifically, Gooding et al [50] compared the effect of a cognitive training game that included empirically validated motivational teaching and rehabilitation techniques (BrainFitnessPlus) with 2 other games: the same previous game without the aforementioned techniques (BrainFitness) and commercially available computer games and puzzles (ie, Brain Age, Sudoku, and crossword puzzles). The study found a statistically significant difference in memory between the groups, favoring BrainFitnessPlus and BrainFitness over commercially available computer games as measured by the Buschke Selective Reminding Test-Delay (BSRT-Delay) and the Wechsler Memory Scale Third Edition-Logical Memory II (WMS-III-LM-II) and favoring BrainFitness over commercially available computer games as measured by the BSRT-Delay only. However, there was no significant difference in memory between the BrainFitnessPlus and BrainFitness groups as measured by the BSRT-Delay and the WMS-III-LM-II [50].

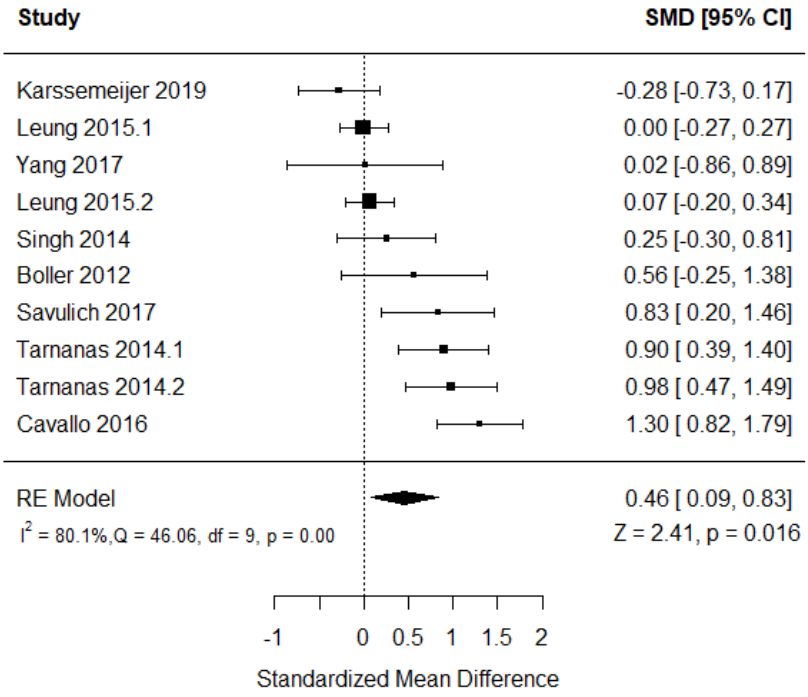
The second trial compared the effect of a cognitive training game with that of exergames [53]. The study found no statistically significant difference (P=.76) in memory between the groups. The last study in this group compared the effect of a cognitive training game that adjusts the level of difficulty of the tasks based on an individual's mastery on each level (ie, adaptive game) with the same game but without adjustment of the level of difficulty of the tasks (ie, nonadaptive game) [58]. The study showed no statistically significant difference between the groups as measured by the WMS-III-LM-II (P=.76) and the California Verbal Learning Test Total Hits (P=.30), but there was a statistically significant difference between the groups as measured by the California Verbal Learning Test II Long Delay Free Recall (P=.03), favoring the adaptive game over the nonadaptive game [58].

Nonverbal Memory

Serious Games Versus Control

The effect of serious games on nonverbal memory was compared with that of no or passive interventions in 44% (8/18) of the studies [49,54-57,60-62]. Of these 8 studies, 2 (25%) assessed nonverbal memory using 2 different measures [55,57]. Therefore, we included the results of all these measures in the meta-analysis to form 9 comparisons (Figure 6) [49,54-57,60-62]. The meta-analysis showed a statistically significant difference (P=.02) in nonverbal memory between the groups, favoring serious games over no or passive interventions (SMD=0.46, 95% CI 0.09-0.83). This difference was also clinically important as the overall effect was outside MCID boundaries (−0.23 to 0.23) and its CI did not cross the "no effect" line (zero effect). For this outcome, the MCID boundaries were calculated as –0.5 times to +0.5 times the SMD value (0.46). The statistical heterogeneity of the evidence was substantial (P<.001; I^2=80.1%). The high heterogeneity may be attributed to differences in sample sizes, participants' health conditions, duration of the intervention, period of the intervention, platform of the intervention, and outcome measures among the studies included in this analysis. The quality of the evidence was very low as it was downgraded by 5 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 8). Subgroup analyses showed no statistically significant difference for all characteristics of the population and intervention (P>.05; Multimedia Appendix 9).

Figure 6. Forest plot of 8 studies (10 comparisons) comparing the effect of serious games with that of control on nonverbal memory. RE: random effect; SMD: standardized mean difference [51,56-59,62-64].

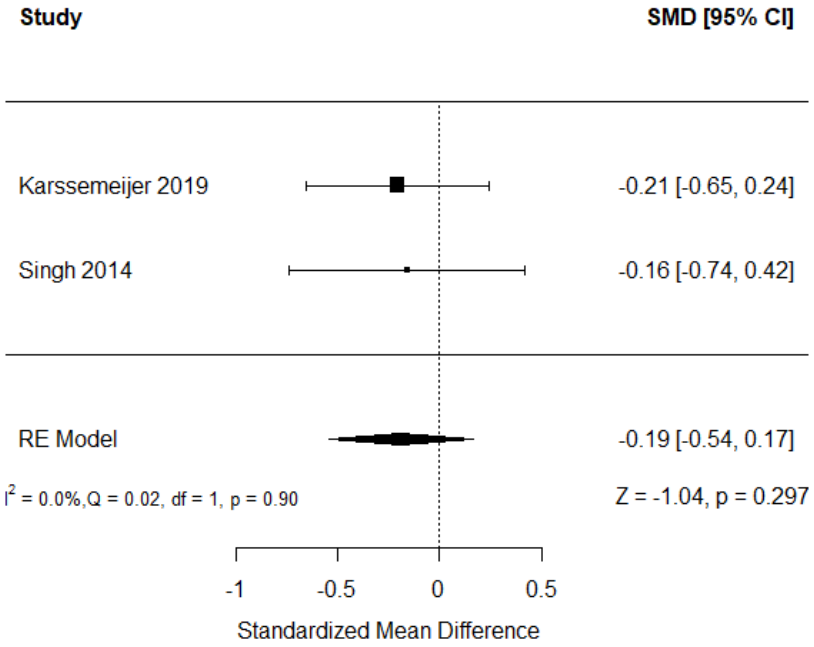


XSL•FO
RenderX

Serious Games Versus Conventional Exercises

The effect of serious games on nonverbal memory was compared with that of conventional exercises in 11% (2/18) of the studies [49,62]. As shown in Figure 7 [49,62], there was no statistically significant difference (P=.30) in nonverbal memory between the groups (SMD=−0.19, 95% CI −0.54 to 0.17). The statistical heterogeneity of the evidence was not a concern (P=.90; I^2=0%). The quality of the evidence was very low as it was downgraded by 3 levels owing to a high risk of bias and imprecision (Multimedia Appendix 8).

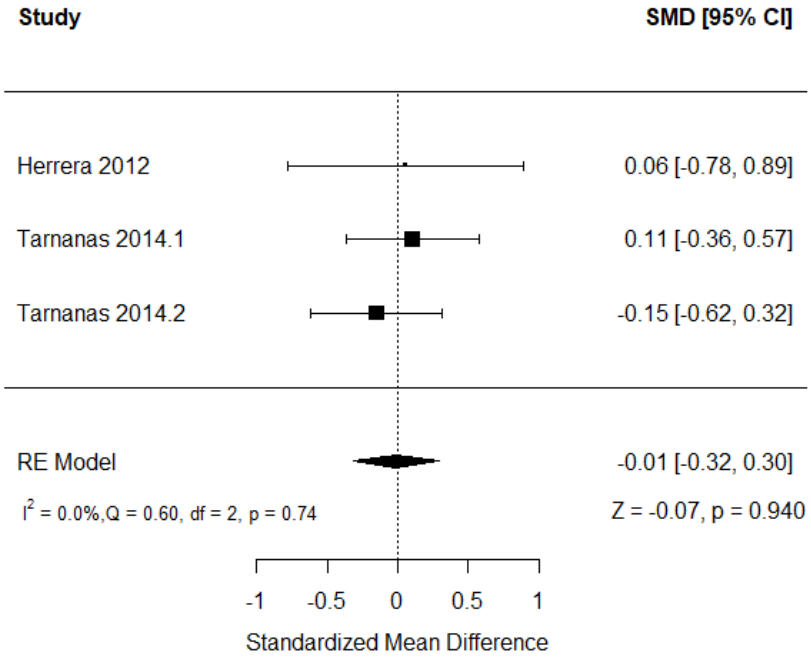
Figure 7. Forest plot of 2 studies comparing the effect of serious games with that of conventional exercises on nonverbal memory. RE: random effect; SMD: standardized mean difference [51,64].



Serious Games Versus Conventional Cognitive Activities

The effect of serious games on nonverbal memory was compared with that of conventional cognitive activities in 11% (2/18) of the studies [57,59]. Of these 2 studies, 1 (50%) assessed nonverbal memory using 2 different measures [57]. Therefore, we included the results of all these measures in the meta-analysis to form 3 comparisons (Figure 8 [57,59]). The meta-analysis showed no statistically significant difference (P=.94) in nonverbal memory between the groups (SMD=−0.01, 95% CI −0.32 to 0.30). The statistical heterogeneity of the evidence was not a concern (P=.74; I^2=0%). The quality of the evidence was very low as it was downgraded by 4 levels owing to a high risk of bias and imprecision (Multimedia Appendix 8).

Figure 8. Forest plot of 2 studies (3 comparisons) comparing the effect of serious games with that of conventional cognitive activities on nonverbal memory. RE: random effect; SMD: standardized mean difference [59,61].



XSL•FO
RenderX

Serious Games Versus Other Serious Games

In total, 17% (3/18) of the studies compared the effect of serious games on nonverbal memory with that of other serious games [50,58,61]. Specifically, Gooding et al [50] compared the effect of BrainFitnessPlus with that of BrainFitness and commercially available computer games. The study showed no statistically significant difference in memory between any 2 of these groups [50].

The second study compared the effect of an adaptive serious game with that of a nonadaptive serious game [58]. The study showed no statistically significant difference in nonverbal memory between the groups as measured by the Rey-Osterrieth complex figure test-delayed recall (P=.25) and the Wechsler Memory Scale Third Edition-Faces II (P=.61) [58].

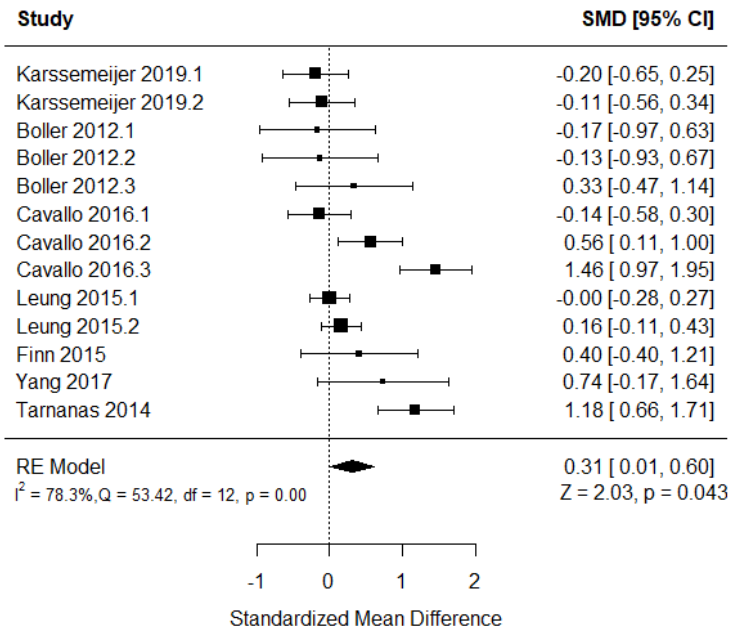
The last study in this group assessed the effect of 2 cognitive training games [61]. Both games consisted of a study and a test phase. In each session of the study phase, both games asked participants to read and remember 16 words presented one at a time on a computer screen for 3 seconds followed by a 1-second white screen [61]. In the test phase, participants were asked to recognize the 16 study words, which were mixed with 16 new words in the first game (recollection training game) and 32 new words in the second game (recognition practice game) [61]. The study showed no statistically significant difference (P=.17) in nonverbal memory between the 2 groups [61].

Working Memory

Serious Games Versus Control

The effect of serious games on working memory was compared with that of control (no or passive interventions) in 39% (7/18) of the studies [52,54-57,61,62]. Of these 7 studies, 4 (57%) assessed working memory using more than one measure [54,55,61,62]. Therefore, we included the results of all these measures in the meta-analysis to form 13 comparisons (Figure 9) [52,54-57,61,62]. The meta-analysis showed a statistically significant difference (P=.04) in working memory between the groups, favoring serious games over no or passive interventions (SMD=0.31, 95% CI 0.01-0.60). This difference was also clinically important as the overall effect was outside MCID boundaries (−0.16 to 0.16) and its CI did not cross the “no effect” line (zero effect). For this outcome, the MCID boundaries were calculated as –0.5 times to +0.5 times the SMD value (0.31). The statistical heterogeneity of the evidence was substantial (P<.001; I^2=78.3%). The high heterogeneity may be attributed to differences in sample sizes, percentage of men, participants’ health conditions, duration of the intervention, period of the intervention, and outcome measures among the studies included in this analysis. The quality of the evidence was very low as it was downgraded by 5 levels owing to a high risk of bias, heterogeneity, and imprecision (Multimedia Appendix 10). Subgroup analyses showed no statistically significant difference for all characteristics of the population and intervention (P>.05; Multimedia Appendix 11).

Figure 9. Forest plot of 7 studies (13 comparisons) comparing the effect of serious games with that of control on working memory. RE: random effect; SMD: standardized mean difference [54,56-59,63,64].

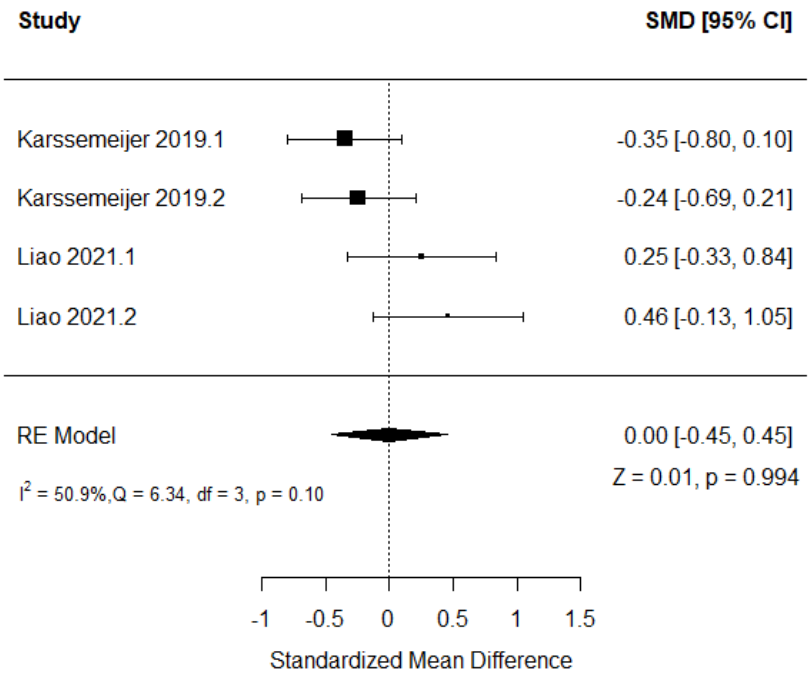


Serious Games Versus Conventional Exercises

The effect of serious games on working memory was compared with that of conventional exercise in 11% (2/18) of the studies [51,62]. Both studies assessed working memory using 2 different measures. Thus, we included the results of all these measures. As shown in Figure 10 [51,62], there was no statistically significant difference (P=.99) in working memory between the serious game and conventional exercise groups (SMD=0.00, 95% CI −0.45 to 0.45). The statistical heterogeneity of the evidence was moderate (P=.10; I^2=50.9%). The quality of the evidence was very low as it was downgraded by 5 levels owing to a high risk of bias and imprecision (Multimedia Appendix 10).

XSL•FO
RenderX

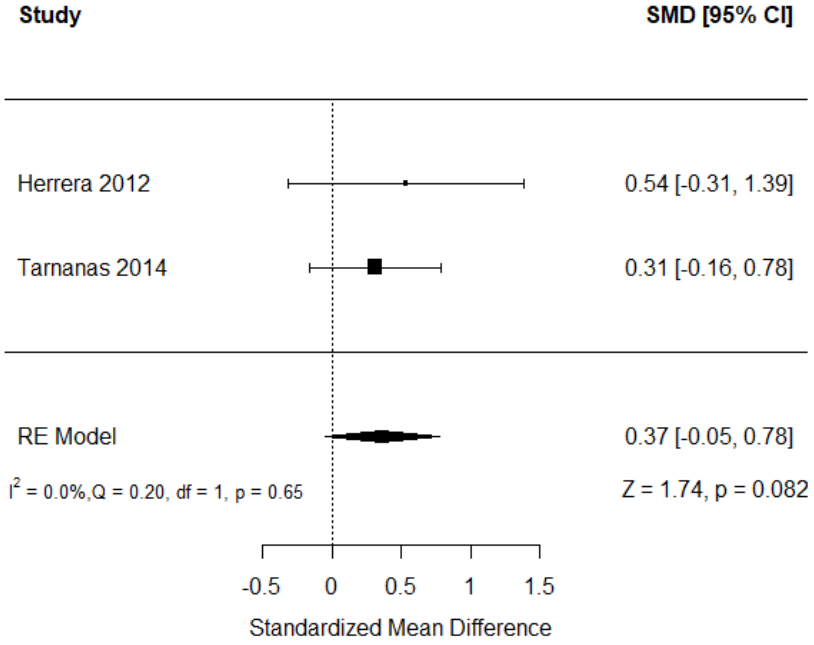
Figure 10. Forest plot of 2 studies (4 comparisons) comparing the effect of serious games with that of conventional exercises on working memory. RE: random effect; SMD: standardized mean difference [53,64].



Serious Games Versus Conventional Cognitive Activities

The effect of serious games on working memory was compared with that of conventional cognitive activities in 11% (2/18) of the studies [57,59] (Figure 11 [57,59]). A meta-analysis of the results of these studies showed no statistically significant difference (P=.08) in working memory between the groups (SMD=0.37, 95% CI −0.05 to 0.78). The statistical heterogeneity of the evidence was not a concern (P=.65; I^2=0%). The quality of the evidence was very low as it was downgraded by 3 levels owing to a high risk of bias and imprecision (Multimedia Appendix 10).

Figure 11. Forest plot of 2 studies comparing the effect of serious games with that of conventional cognitive activities on working memory. RE: random effect; SMD: standardized mean difference [59,61].



Serious Games Versus Other Serious Games

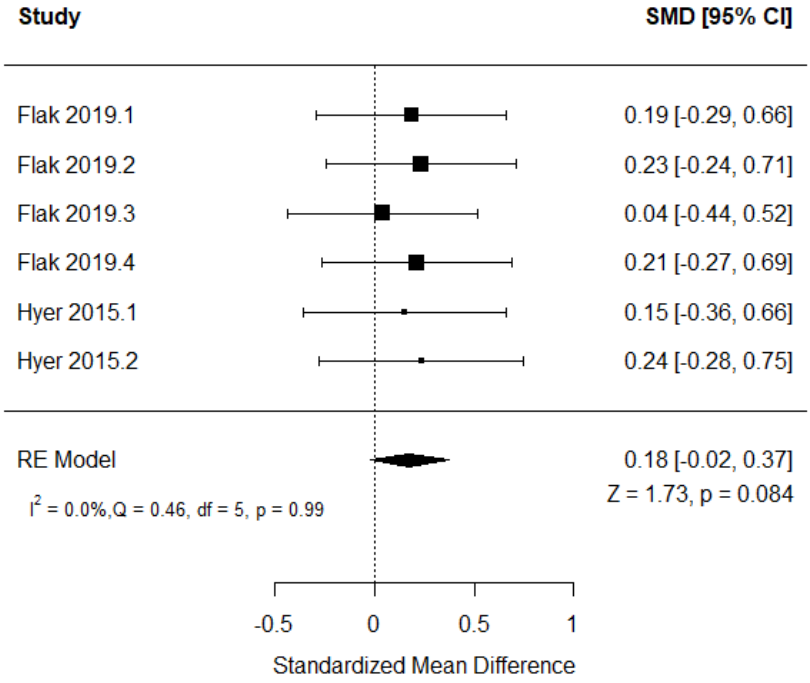
The effect of serious games on working memory was compared with that of other serious games in 22% (4/18) of the studies [53,58,61,63]. Specifically, the first study compared the effect of a cognitive training game with that of exergames [53]. The study found a statistically significant difference (P<.001) in memory between the groups, favoring cognitive training games over exergames [53]. The second study assessed the effect of 2 cognitive training games on working memory: a recollection

XSL•FO
RenderX

training game and a recognition practice game [61]. The study showed no statistically significant difference in working memory between the 2 groups as measured by the n-back task (P=.78) and reading span task (P=.76) [61].

The remaining 50% (2/4) of the studies compared the effect of adaptive serious games with that of nonadaptive serious games [58,63]. Of the 2 studies, 1 (50%) assessed working memory using 4 different measures [58], whereas the other study (50%) used 2 different measures to do so [63]. Hence, we included the results of all these measures in the meta-analysis to form 6 comparisons. As shown in Figure 12 [58,63], there was no statistically significant difference (P=.08) in working memory between adaptive serious games and nonadaptive serious games (SMD=0.18, 95% CI −0.02 to 0.37). The statistical heterogeneity of the evidence was not a concern (P=.99; I^2=0%). The quality of the evidence was low as it was downgraded by 2 levels owing to a high risk of bias and imprecision (Multimedia Appendix 10).

Figure 12. Forest plot of 2 studies (6 comparisons) comparing the effect of adaptive serious games with that of nonadaptive serious games on working memory. RE: random effect; SMD: standardized mean difference [60,65].



Discussion

Principal Findings

This study summarized the evidence regarding the effectiveness of serious games in improving memory. Our meta-analyses showed that serious games are more effective than no or passive interventions in improving nonverbal and working memory. Surprisingly, we found that serious games are as effective as no or passive interventions in improving verbal memory, which, therefore, deems serious games ineffective. This review demonstrated that serious games are more effective than conventional exercises in improving verbal memory. However, we found that serious games are as effective as conventional exercises in improving nonverbal and working memory, indicating that serious games are comparable with conventional exercises. Evidence suggests that cognitive training and exercise work through distinct neuronal mechanisms and, therefore, if combined, they might have synergistic and more effective results compared with being used as separate interventions [65,66]. Studying this synergistic relationship will become important in future primary research and trials. With the advances in virtual reality technologies, their availability, and rising applications of the metaverse [67], more evidence is needed to assess the effectiveness of virtual reality–based exergames in improving memory [68].

The meta-analyses in this review showed that serious games are as effective as conventional cognitive training in improving verbal, nonverbal, and working memory, meaning that serious games and conventional cognitive training are comparable. Furthermore, we found that the effect of adaptive serious games is similar to that of nonadaptive serious games in improving working memory.

The findings of our review and previous reviews were consistent for some outcomes and different for others. Specifically, a systematic review conducted by Lampit et al [32] compared the effect of cognitive training games with that of passive and active interventions on verbal, nonverbal, and working memory in healthy older adults. Consistent with our findings, the review found no statistically significant difference (P>.05) in the effect of cognitive training games and of no or passive interventions on verbal memory, and there was a statistically significant difference in working memory between the groups, favoring cognitive training games over no or passive interventions [32]. In contrast to our findings, Lampit et al [32] did not find a statistically significant difference in nonverbal memory between the groups. The contrary finding may be attributed to the following reasons: (1) although the number of participants was

XSL•FO
RenderX

≥100 in 17% (3/18) of the studies in our meta-analysis, the number of participants was <100 in all studies included in the meta-analysis by Lampit et al [32]; (2) the total number of training hours was >15 in only 2 studies included in the meta-analysis by Lampit et al [32], whereas the total number of training hours was >15 in 28% (5/18) of the studies in our meta-analysis; and (3) all studies meta-analyzed in our review (15/18, 83%) recruited participants with cognitive impairment, whereas all studies meta-analyzed in the review by Lampit et al [32] recruited participants without cognitive impairment.

In contrast to our findings, Lampit et al [32] found a statistically significant difference in verbal, nonverbal, and working memory between the groups, favoring cognitive training games over active interventions [32]. This may be attributed to the following reasons: (1) our findings related to these comparisons are based on meta-analyses of 11% (2/18) of the studies, whereas the findings of Lampit et al [32] are based on meta-analyses of 6 to 15 studies; (2) Lampit et al [32] compared cognitive training games with all active interventions, but this review compared cognitive training games with specific active interventions (ie, conventional exercises and conventional cognitive training); (3) all studies meta-analyzed in our review (15/18, 83%) recruited participants with cognitive impairment, whereas all studies meta-analyzed in the review by Lampit et al [32] recruited participants without cognitive impairment; and (4) the review by Lampit et al [32] included some pilot RCTs, whereas our review included only RCTs.

Another review examined the effect of cognitive training games on verbal and working memory among healthy older adults regardless of the comparator type (ie, passive and active controls) [30]. Meta-analyses in that review showed a statistically significant difference in verbal memory (P=.03) and working memory (P<.001) between the groups, favoring cognitive training games over all types of comparators [30].

Hill et al [34] conducted a systematic review to assess the effect of cognitive training games on verbal, nonverbal, and working memory among people with MCI or dementia regardless of the comparator type. For people with MCI, the review found a statistically significant difference between the groups in verbal and working memory (P<.001), favoring all types of comparators [34]. In contrast, there was no statistically significant effect of cognitive training games on nonverbal memory when compared with all types of comparators [34]. For people with dementia, the review showed no statistically significant effect of cognitive training games on verbal, nonverbal, or working memory when compared with all types of comparators [34].

Research and Practical Implications

Research Implications

Given that the review focused on memory among older adults with cognitive impairment, future reviews should assess the effectiveness of serious games on other cognitive functions (eg, learning, language, executive function, and processing speed) in young and older adults with or without cognitive impairment. In this review, a few studies (≤3) were included in the meta-analyses that compared serious games with active interventions (ie, conventional exercises, conventional cognitive training, and nonadaptive serious games); therefore, our findings regarding these comparisons remain inconclusive. Thus, there is a pressing need to conduct more studies to compare the effect of serious games on memory with active interventions.

Most studies in this review (12/18, 67%) were carried out in clinical settings, thus offering the researchers more control over the experiments. However, the participants may have been stressed by playing these games outside the environment that they were used to. Therefore, more studies should be conducted in the community and home settings, allowing participants to be at ease and enabling the researchers to examine other factors that could come into play, such as environmental conditions (eg, room temperature and lighting).

In this review, the long-term effect of serious games was not assessed as few studies reported follow-up data, and the follow-up period was not consistent among the studies. Further studies should assess the long-term effect of serious games on memory. Most of the included studies (15/18, 83%) did not report the mean and SD of pre-post intervention change in memory for each group. Researchers should report this information to accurately calculate effect sizes.

Future studies should also examine and compare the effectiveness of playing serious games in multiplayer mode with other members of the family or community as this has not been assessed in previous studies. We urge researchers to conduct and report RCTs following recommended guidelines or tools (eg, RoB 2 [39]) to avoid the biases identified in this review.

Practical Implications

This review shows that serious games can be effective in improving verbal, nonverbal, and working memory. However, these findings should be interpreted cautiously given that most meta-analyses were based on a few studies (≤3) and judged to have a low quality of evidence for the following reasons: most of the included studies (11/18, 61%) were judged to have some concerns regarding the overall bias, the heterogeneity of the evidence was high in approximately half of the meta-analyses (4/10, 40%), and the total effect sizes were imprecise in all meta-analyses (10/10, 100%). On the basis of our review findings, serious games are still not ready as substitutes for real-world interactions and experiences; they should still be used as a supplement rather than an alternative method for interventions targeting the improvement of verbal, nonverbal, and working memory until more evidence suggests otherwise.

Despite the ubiquity and availability of smart mobile devices (ie, tablets and smartphones), only 6% (1/18) of the included studies used them [60]. Mobile devices can be more pervasive and accessible than PCs or commercially available gaming consoles. Studies estimate that, in 2021 alone, approximately 15 billion mobile devices exist worldwide and are used by >7.1 billion users [69]; this is expected to rise. Game and app developers should invest in creating serious games on mobile devices that target improving verbal, nonverbal, and working memory.

XSL•FO
RenderX

Limitations

This review cannot comment on the effectiveness of serious games (1) delivered on nondigital platforms, (2) used for other purposes (eg, screening or diagnosis), (3) used for improving other cognitive abilities (eg, learning, processing speed, and executive functions), (4) among other age groups, or (5) among those without cognitive impairment. This is because such interventions, outcomes, and populations were beyond the scope of this review.

It is likely that we missed some relevant studies as this review did not search some databases (eg, PubMed and the Cochrane Library [CENTRAL]) and excluded studies that were quasi-experiments, pilot RCTs, published before 2010, or written in non-English languages. The quality of the evidence was very low in all meta-analyses except for 10% (1/10); this may decrease the internal validity of our findings. We cannot comment on the long-term effect of serious games on memory as this review focused on the short-term effect of serious games by meta-analyzing only postintervention data rather than follow-up data. This is because the follow-up period was not consistent among the studies.

The effect size for each meta-analyzed study was likely overestimated or underestimated in this review given that the authors used postintervention data for each group to assess the effect size rather than the pre-post intervention change for each group. Postintervention outcome data were used as most studies (15/18, 83%) did not report the mean and SD for pre-post intervention change in memory for each group, and there was no statistically significant difference in memory between the groups at baseline in all studies (18/18, 100%).

Conclusions

Serious games may have a significant role to play in improving verbal, nonverbal, and working memory in older adults with cognitive impairment. However, these findings should be treated with caution given that most meta-analyses (7/10, 70%) were based on a few studies (≤3) and judged to have a low quality of evidence for the following reasons: most of the included studies (11/18, 61%) were judged to have some concerns regarding the overall bias, the heterogeneity of the evidence was high in approximately half of the meta-analyses (4/10, 40%), and the total effect sizes were imprecise in all meta-analyses (10/10, 100%). Therefore, serious games should be offered as a supplement to existing proven and safe interventions rather than as a complete substitute until further, more robust evidence is available. Further reviews are necessary to investigate the short- and long-term effect of serious games on memory and other cognitive abilities (eg, executive function, processing speed, and learning) among people of different age groups with or without cognitive impairment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 26 KB - games_v10i3e35202_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 33 KB - games_v10i3e35202_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 25 KB - games_v10i3e35202_app3.docx]

Multimedia Appendix 4
Excluded studies.
[XLSX File (Microsoft Excel File), 99 KB - games_v10i3e35202_app4.xlsx]

Multimedia Appendix 5
Reviewers' judgments about each risk of bias domain for each included study.
[DOCX File , 92 KB - games_v10i3e35202_app5.docx]

Multimedia Appendix 6
Grading of Recommendations Assessment, Development, and Evaluation profile for comparison of serious games with control, conventional exercises, and conventional cognitive activities for verbal memory.
[DOCX File , 17 KB - games_v10i3e35202_app6.docx]

Multimedia Appendix 7

XSL•FO
RenderX

Moderation analyses for verbal memory.
[DOCX File , 17 KB - games_v10i3e35202_app7.docx]

Multimedia Appendix 8
Grading of Recommendations Assessment, Development, and Evaluation profile for comparison of serious games with control, conventional exercises, conventional cognitive activities, and other serious games for nonverbal memory.
[DOCX File , 17 KB - games_v10i3e35202_app8.docx]

Multimedia Appendix 9
Moderation analyses for nonverbal memory.
[DOCX File , 17 KB - games_v10i3e35202_app9.docx]

Multimedia Appendix 10
Grading of Recommendations Assessment, Development, and Evaluation profile for comparison of serious games with control, conventional exercises, conventional cognitive activities, and other serious games for working memory.
[DOCX File , 18 KB - games_v10i3e35202_app10.docx]

Multimedia Appendix 11
Moderation analyses for working memory.
[DOCX File , 17 KB - games_v10i3e35202_app11.docx]

References

1. Hao L, Xu X, Dupre ME, Guo A, Zhang X, Qiu L, et al. Adequate access to healthcare and added life expectancy among older adults in China. BMC Geriatr 2020 Apr 09;20(1):129 [FREE Full text] [doi: 10.1186/s12877-020-01524-9] [Medline: 32272883]
2. Maldonado Briegas JJ, Sánchez Iglesias AI, Ballester SG, Vicente Castro F. The well-being of the elderly: memory and aging. Front Psychol 2020 May 26;11:778 [FREE Full text] [doi: 10.3389/fpsyg.2020.00778] [Medline: 32528338]
3. World Population Ageing 2019. Department of Economic and Social Affairs, Population Division, United Nations. 2020. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Report.pdf [accessed 2021-11-30]
4. UN Report: World Population Ageing 2020: Highlights. United Nations. 2020 Dec 19. URL: https://www.thehinducentre.com/incoming/article33347255.ece#:~:text=All%20regions%20will%20see%20an,16.0%20per%20cent%20in%202050 [accessed 2021-11-30]
5. Al Mahmud A, Slikboer R, Stargatt J, Bhar S. Computer-based cognitive interventions for mild cognitive impairment and dementia in older adults: protocol for a systematic review of published studies and meta-analysis. Syst Rev 2019 Sep 06;8(1):231 [FREE Full text] [doi: 10.1186/s13643-019-1146-x] [Medline: 31492180]
6. Qin HY, Zhao XD, Zhu BG, Hu CP. Demographic factors and cognitive function assessments associated with mild cognitive impairment progression for the elderly. Biomed Res Int 2020 Feb 8;2020:3054373 [FREE Full text] [doi: 10.1155/2020/3054373] [Medline: 32090075]
7. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
8. World population prospects: the 2015 revision, key findings and advance tables. Department of Economics and Social Affairs, Population Division, United Nations. 2015. URL: https://population.un.org/wpp/Publications/Files/Key_Findings_WPP_2015.pdf [accessed 2021-11-30]
9. Mild Cognitive Impairment (MCI). Alzheimer's Association. 2021. URL: https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment [accessed 2021-02-05]
10. Sanford AM. Mild cognitive impairment. Clin Geriatr Med 2017 Aug;33(3):325-337. [doi: 10.1016/j.cger.2017.02.005] [Medline: 28689566]
11. Jongsiriyanyong S, Limpawattana P. Mild cognitive impairment in clinical practice: a review article. Am J Alzheimers Dis Other Demen 2018 Dec;33(8):500-507 [FREE Full text] [doi: 10.1177/1533317518791401] [Medline: 30068225]
12. APA Dictionary of Psychology-verbal memory. American Psychological Association. 2020. URL: https://dictionary.apa.org/verbal-memory [accessed 2021-11-23]
13. Non-Verbal Memory-Neuropsychology-Cognitive Ability. Cognifit. 2021. URL: https://www.cognifit.com/science/cognitive-skills/nonverbal-memory [accessed 2021-11-30]
14. APA Dictionary of Psychology- Short-term memory (STM). American Psychological Association. 2020. URL: https://dictionary.apa.org/short-term-memory [accessed 2021-11-23]
15. APA Dictionary of Psychology- Long-term memory (LTM). American Psychological Association. 2020. URL: https://dictionary.apa.org/long-term-memory [accessed 2021-11-23]

XSL•FO
RenderX

16. Cowan N. Working memory underpins cognitive development, learning, and education. Educ Psychol Rev 2014 Jun 01;26(2):197-223 [FREE Full text] [doi: 10.1007/s10648-013-9246-y] [Medline: 25346585]
17. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
18. Manera V, Ben-Sadoun G, Aalbers T, Agopyan H, Askenazy F, Benoit M, et al. Recommendations for the use of serious games in neurodegenerative disorders: 2016 Delphi panel. Front Psychol 2017 Jun 25;8:1243 [FREE Full text] [doi: 10.3389/fpsyg.2017.01243] [Medline: 28790945]
19. Muñoz JE, Montoya MF, Boger J. Chapter 7 From exergames to immersive virtual reality systems: serious games for supporting older adults. In: Choukou MA, Syed-Abdul S, editors. Smart Home Technologies and Services for Geriatric Rehabilitation. Cambridge, MA, USA: Academic Press; 2021:141-204.
20. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020 Jun 30;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
21. Bonnechère B, Omelina L, Jansen B, Van Sint Jan S. Balance improvement after physical therapy training using specially developed serious games for cerebral palsy children: preliminary results. Disabil Rehabil 2017 Feb;39(4):403-406. [doi: 10.3109/09638288.2015.1073373] [Medline: 28033958]
22. Meijer HA, Graafland M, Goslings JC, Schijven MP. Systematic review on the effects of serious games and wearable technology used in rehabilitation of patients with traumatic bone and soft tissue injuries. Arch Phys Med Rehabil 2018 Sep;99(9):1890-1899. [doi: 10.1016/j.apmr.2017.10.018] [Medline: 29138050]
23. Abd-Alrazaq A, Al-Jafar E, Alajlani M, Toro C, Alhuwail D, Ahmed A, et al. The effectiveness of serious games for alleviating depression: systematic review and meta-analysis. JMIR Serious Games 2022 Jan 14;10(1):e32331 [FREE Full text] [doi: 10.2196/32331] [Medline: 35029530]
24. Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Akhu-Zaheya L, Ahmed A, et al. The effectiveness of serious games in alleviating anxiety: systematic review and meta-analysis. JMIR Serious Games 2022 Feb 14;10(1):e29137 [FREE Full text] [doi: 10.2196/29137] [Medline: 35156932]
25. Abd-Alrazaq A, Alajlani M, Alhuwail D, Toro C, Giannicchi A, Ahmed A, et al. The effectiveness and safety of serious games for improving cognitive abilities among elderly people with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Mar 10;10(1):e34592 [FREE Full text] [doi: 10.2196/34592] [Medline: 35266877]
26. Jones CM, Scholes L, Johnson D, Katsikitis M, Carras MC. Gaming well: links between videogames and flourishing mental health. Front Psychol 2014 Mar 31;5:260 [FREE Full text] [doi: 10.3389/fpsyg.2014.00260] [Medline: 24744743]
27. Pallavicini F, Ferrari A, Mantovani F. Video games for well-being: a systematic review on the application of computer games for cognitive and emotional training in the adult population. Front Psychol 2018 Nov 7;9:2127 [FREE Full text] [doi: 10.3389/fpsyg.2018.02127] [Medline: 30464753]
28. Sokolov AA, Collignon A, Bieler-Aeschlimann M. Serious video games and virtual reality for prevention and neurorehabilitation of cognitive decline because of aging and neurodegeneration. Curr Opin Neurol 2020 Apr;33(2):239-248. [doi: 10.1097/WCO.0000000000000791] [Medline: 32073439]
29. Stern Y, Blumen HM, Rich LW, Richards A, Herzberg G, Gopher D. Space Fortress game training and executive control in older adults: a pilot intervention. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2011 Nov;18(6):653-677 [FREE Full text] [doi: 10.1080/13825585.2011.613450] [Medline: 21988726]
30. Bonnechère B, Langley C, Sahakian BJ. The use of commercial computerised cognitive games in older adults: a meta-analysis. Sci Rep 2020 Sep 17;10(1):15276 [FREE Full text] [doi: 10.1038/s41598-020-72281-3] [Medline: 32943742]
31. Wang G, Zhao M, Yang F, Cheng LJ, Lau Y. Game-based brain training for improving cognitive function in community-dwelling older adults: a systematic review and meta-regression. Arch Gerontol Geriatr 2021;92:104260. [doi: 10.1016/j.archger.2020.104260] [Medline: 32980574]
32. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
33. Yen HY, Chiu HL. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
34. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
35. Abd-Alrazaq A, Alhuwail D, Ahmed A, Househ M. Effectiveness of serious games for improving executive functions among older adults with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Jul 25;10(3):e36123 [FREE Full text] [doi: 10.2196/36123] [Medline: 35877166]

XSL•FO
RenderX

36. Abd-Alrazaq A, Ahmed A, Alali H, Aldardour AM, Househ M. The effectiveness of serious games on the cognitive processing speed among elderly people with cognitive impairment: a systematic review and meta-analysis. JMIR Serious Games (forthcoming) 2022 Apr 7:e36754 [FREE Full text] [doi: 10.2196/36754]
37. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
38. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977 Jun;33(2):363-374. [Medline: 884196]
39. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
40. Fernández-Castilla B, Jamshidi L, Declercq L, Beretvas SN, Onghena P, Van den Noortgate W. The application of meta-analytic (multi-level) models with multiple random effects: a systematic review. Behav Res Methods 2020 Oct;52(5):2031-2052. [doi: 10.3758/s13428-020-01373-9] [Medline: 32162276]
41. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. Res Synth Methods 2011 Mar;2(1):61-76. [doi: 10.1002/jrsm.35] [Medline: 26061600]
42. Nakagawa S, Santos E. Methodological issues and advances in biological meta-analysis. Evol Ecol 2012 Feb 17;26(5):1253-1274. [doi: 10.1007/s10682-012-9555-5]
43. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010 Aug 5;36(3):1-48. [doi: 10.18637/jss.v036.i03]
44. Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Sussex, UK: John Wiley & Sons; 2008:243-296.
45. Schunemann HJ, Oxman AD, Vist GE, Higgins JP, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Sussex, UK: John Wiley & Sons; 2008:359-387.
46. Valdes EG, O'Connor ML, Edwards JD. The effects of cognitive speed of processing training among older adults with psychometrically- defined mild cognitive impairment. Curr Alzheimer Res 2012 Nov;9(9):999-1009. [doi: 10.2174/156720512803568984] [Medline: 22594383]
47. Zhuang JP, Fang R, Feng X, Xu XH, Liu LH, Bai QK, et al. The impact of human-computer interaction-based comprehensive training on the cognitive functions of cognitive impairment elderly individuals in a nursing home. J Alzheimers Dis 2013;36(2):245-251. [doi: 10.3233/JAD-130158] [Medline: 23587747]
48. Hagovská M, Takáč P, Dzvoník O. Effect of a combining cognitive and balanced training on the cognitive, postural and functional status of seniors with a mild cognitive deficit in a randomized, controlled trial. Eur J Phys Rehabil Med 2016 Feb;52(1):101-109 [FREE Full text] [Medline: 26325026]
49. Fiatarone Singh MA, Gates N, Saigal N, Wilson GC, Meiklejohn J, Brodaty H, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. J Am Med Dir Assoc 2014 Dec;15(12):873-880. [doi: 10.1016/j.jamda.2014.09.010] [Medline: 25444575]
50. Gooding AL, Choi J, Fiszdon JM, Wilkins K, Kirwin PD, van Dyck CH, et al. Comparing three methods of computerised cognitive training for older adults with subclinical cognitive decline. Neuropsychol Rehabil 2016 Oct;26(5-6):810-821. [doi: 10.1080/09602011.2015.1118389] [Medline: 26674122]
51. Liao YY, Chen IH, Hsu WC, Tseng HY, Wang RY. Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: a randomised controlled trial. Ann Phys Rehabil Med 2021 Sep;64(5):101492. [doi: 10.1016/j.rehab.2021.101492] [Medline: 33454398]
52. Finn M, McDonald S. Repetition-lag training to improve recollection memory in older people with amnestic mild cognitive impairment. A randomized controlled trial. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2015;22(2):244-258. [doi: 10.1080/13825585.2014.915918] [Medline: 24820545]
53. Park JH, Park JH. Does cognition-specific computer training have better clinical outcomes than non-specific computer training? A single-blind, randomized controlled trial. Clin Rehabil 2018 Feb;32(2):213-222. [doi: 10.1177/0269215517719951] [Medline: 28726492]
54. Cavallo M, Hunter EM, van der Hiele K, Angilletta C. Computerized structured cognitive training in patients affected by early-stage Alzheimer's disease is feasible and effective: a randomized controlled study. Arch Clin Neuropsychol 2016 Dec 01;31(8):868-876. [doi: 10.1093/arclin/acw072] [Medline: 27600448]
55. Leung NT, Tam HM, Chu LW, Kwok TC, Chan F, Lam LC, et al. Neural plastic effects of cognitive training on aging brain. Neural Plast 2015;2015:535618 [FREE Full text] [doi: 10.1155/2015/535618] [Medline: 26417460]
56. Yang Y, Kwak YT. Improvement of cognitive function after computer-based cognitive training in early stage of Alzheimer's dementia. Dement Neurocogn Disord 2017 Mar;16(1):7-11 [FREE Full text] [doi: 10.12779/dnd.2017.16.1.7] [Medline: 30906364]

XSL•FO
RenderX

57. Tarnanas I, Tsolakis A, Tsolaki M. Assessing virtual reality environments as cognitive stimulation method for patients with MCI. In: Brooks AL, Brahnam S, Jain LC, editors. Technologies of Inclusive Well-Being: Serious Games, Alternative Realities, and Play Therapy. Berlin, Germany: Springer; 2014:39-74.
58. Flak MM, Hol HR, Hernes SS, Chang L, Engvig A, Bjuland KJ, et al. Adaptive computerized working memory training in patients with mild cognitive impairment. A randomized double-blind active controlled trial. Front Psychol 2019 Apr 12;10:807 [FREE Full text] [doi: 10.3389/fpsyg.2019.00807] [Medline: 31031677]
59. Herrera C, Chambon C, Michel BF, Paban V, Alescio-Lautier B. Positive effects of computer-based cognitive training in adults with mild cognitive impairment. Neuropsychologia 2012 Jul;50(8):1871-1881. [doi: 10.1016/j.neuropsychologia.2012.04.012] [Medline: 22525705]
60. Savulich G, Piercy T, Fox C, Suckling J, Rowe JB, O'Brien JT, et al. Cognitive training using a novel memory game on an iPad in patients with amnestic mild cognitive impairment (aMCI). Int J Neuropsychopharmacol 2017 Aug 01;20(8):624-633 [FREE Full text] [doi: 10.1093/ijnp/pyx040] [Medline: 28898959]
61. Boller B, Jennings JM, Dieudonné B, Verny M, Ergis AM. Recollection training and transfer effects in Alzheimer's disease: effectiveness of the repetition-lag procedure. Brain Cogn 2012 Mar;78(2):169-177. [doi: 10.1016/j.bandc.2011.10.011] [Medline: 22122949]
62. Karssemeijer EG, Aaronson JA, Bossers WJ, Donders R, Olde Rikkert MG, Kessels RP. The quest for synergy between physical exercise and cognitive stimulation via exergaming in people with dementia: a randomized controlled trial. Alzheimers Res Ther 2019 Jan 05;11(1):3 [FREE Full text] [doi: 10.1186/s13195-018-0454-z] [Medline: 30611286]
63. Hyer L, Scott C, Atkinson MM, Mullen CM, Lee A, Johnson A, et al. Cognitive training program to improve working memory in older adults with MCI. Clin Gerontol 2016;39(5):410-427. [doi: 10.1080/07317115.2015.1120257] [Medline: 29471774]
64. Borenstein M, Higgins JP. Meta-analysis and subgroups. Prev Sci 2013 Apr;14(2):134-143. [doi: 10.1007/s11121-013-0377-7] [Medline: 23479191]
65. Salisbury D, Plocher T, Yu F. Efficacy of simultaneous aerobic exercise and cognitive training in subjective cognitive decline: study protocol for randomized controlled trial of the Exergames Study. Trials 2021 Jan 06;22(1):14 [FREE Full text] [doi: 10.1186/s13063-020-04950-7] [Medline: 33407727]
66. Suo C, Singh MF, Gates N, Wen W, Sachdev P, Brodaty H, et al. Therapeutically relevant structural and functional mechanisms triggered by physical and cognitive exercise. Mol Psychiatry 2016 Nov;21(11):1633-1642 [FREE Full text] [doi: 10.1038/mp.2016.19] [Medline: 27001615]
67. Kye B, Han N, Kim E, Park Y, Jo S. Educational applications of metaverse: possibilities and limitations. J Educ Eval Health Prof 2021;18:32 [FREE Full text] [doi: 10.3352/jeehp.2021.18.32] [Medline: 34897242]
68. Huang KT. Exergaming executive functions: an immersive virtual reality-based cognitive training for adults aged 50 and older. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):143-149. [doi: 10.1089/cyber.2019.0269] [Medline: 31794673]
69. Mobile statistics report, 2021-2025. The Radicati Group. 2025. URL: https://www.radicati.com/wp/wp-content/uploads/2020/12/Mobile-Statistics-Report-2021-2025-Brochure.pdf [accessed 2021-11-30]

Abbreviations

BSRT-Delay: Buschke Selective Reminding Test-Delay
MCI: mild cognitive impairment
MCID: minimal clinically important difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: Risk of Bias 2
SMD: standardized mean difference
WMS-III-LM-II: Wechsler Memory Scale Third Edition-Logical Memory II

Edited by N Zary; submitted 29.11.21; peer-reviewed by M Alajlani, A Giannicchi, F Velayati, H Ayatollahi, E Karssemeijer, LJ Cheng; comments to author 03.02.22; revised version received 10.02.22; accepted 13.05.22; published 09.08.22.

Please cite as:
Abd-alrazaq A, Alhuwail D, Al-Jafar E, Ahmed A, Shuweihdi F, Reagu SM, Househ M
The Effectiveness of Serious Games in Improving Memory Among Older Adults With Cognitive Impairment: Systematic Review and Meta-analysis
JMIR Serious Games 2022;10(3):e35202
URL: https://games.jmir.org/2022/3/e35202
doi:10.2196/35202
PMID:35943792

XSL•FO
RenderX

©Alaa Abd-alrazaq, Dari Alhuwail, Eiman Al-Jafar, Arfan Ahmed, Farag Shuweihdi, Shuja Mohd Reagu, Mowafa Househ. Originally published in JMIR Serious Games (https://games.jmir.org), 09.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Interventions for Emotion Regulation in Children and Early Adolescents: Systematic Review and Meta-analysis

Sally Reynard[1], MSc; Joao Dias[2,3,4], PhD; Marija Mitic[5], MD; Beate Schrank[5,6], MD, PhD; Kate Anne Woodcock[1,7], PhD

[1]Centre for Applied Psychology, University of Birmingham, Birmingham, United Kingdom

[2]Faculty of Science and Technology, University of Algarve, Faro, Portugal

[3]Algarve Centre of Marine Sciences, Faro, Portugal

[4]Institute of Systems and Computer Engineering: Research and Development, Lisbon, Portugal

[5]Die Offene Tür Research Group for Mental Health of Children and Adolescents, Ludwig Boltzmann Society, Karl Landsteiner University of Health Sciences, Krems, Austria

[6]Department of Psychiatry and Psychotherapy, University Hospital Tulln, Tulln, Austria

[7]Institute of Mental Health, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Sally Reynard, MSc
Centre for Applied Psychology
University of Birmingham
52 Pritchatts Road
Birmingham, B15 2SA
United Kingdom
Phone: 44 0121 414 6036
Email: sallyreynard1@gmail.com

Abstract

Background: Difficulties in emotion regulation are common in adolescence and are associated with poor social and mental health outcomes. However, psychological therapies that promote adaptive emotion regulation may be inaccessible and unattractive to youth. Digital interventions may help address this need.

Objective: The aim of this systematic review and meta-analysis was to synthesize evidence on the efficacy, feasibility, and acceptability of emotion regulation digital interventions in children and early adolescents aged 8 to 14 years.

Methods: Systematic searches of Web of Science, MEDLINE, PsycINFO, EMBASE, Education Resources Information Centre, ACM Digital Library, and IEEE Xplore up to July 2020 identified 39 studies, of which 11 (28%) were included in the meta-analyses (n=2476 participants). A bespoke tool was used to assess risk of bias.

Results: The studies evaluated digital games (27/39, 69%), biofeedback (4/39, 10%), virtual or augmented reality (4/39, 10%), and program or multimedia (4/39, 10%) digital interventions in samples classified as diagnosed, at risk, healthy, and universal. The most consistent evidence came from digital games, which reduced negative emotional experience with a small significant effect, largely in youth at risk of anxiety (Hedges g=–0.19, 95% CI –0.34 to –0.04). In general, digital interventions tended to improve emotion regulation, but this effect was not significant (Hedges g=0.19, 95% CI –0.16 to 0.54).

Conclusions: Most feasibility issues were identified in diagnosed youth, and acceptability was generally high across intervention types and samples. Although there is cause to be optimistic about digital interventions supporting the difficulties that youth experience in emotion regulation, the predominance of early-stage development studies highlights the need for more work in this area.

(JMIR Serious Games 2022;10(3):e31456) doi:10.2196/31456

KEYWORDS

emotion regulation; digital interventions; youth; systematic review; meta-analysis; children; early adolescents; serious games; training; biofeedback; mobile phone

XSL•FO
RenderX

Introduction

Background

Emotion regulation difficulties are prospectively associated with negative social outcomes [1] and psychiatric disorders in youth [2]. This is particularly significant considering that half of all lifetime psychiatric disorders begin by the age of 14 years [3], and a recent large-scale meta-analysis reported 14.5 years as the worldwide peak age of psychiatric disorder onset [4]. The malleability of affective neural circuitry is heightened from late childhood through early adolescence [5,6]. Hence, this is a period of particular interest for harnessing adaptive emotion regulation strategies, which may support positive social outcomes and psychological well-being [7-9]. Proximal social input that manipulates the environment in a positive manner, such as targeted intervention, can improve emotion regulation ability [10,11]. Digital interventions may constitute efficacious, accessible, and attractive interventions in youth [12]. However, there is no systematic understanding of existing emotion regulation digital interventions and their efficacy in youth. Consequently, the aims of this systematic review and meta-analysis were to present a comprehensive understanding of the extant evidence on digital interventions that target emotion regulation in youth to provide recommendations for this emerging field.

Emotion Regulation in Youth

Emotion regulation is operationalized as the attempt to recognize positive and negative emotional reactions in ourselves and to increase or decrease these in ourselves or others [13,14]. The Extended Process Model of emotion regulation provides a framework of emotion regulation stages (identification, selection, implementation, and monitoring) and strategies in relation to an emotional goal (refer to the study by Gross [15] for a comprehensive account) and is consistent with the way many extant digital interventions for emotion regulation have been designed.

The developmental trajectories of improvements in different stages of the Extended Process Model are not equivalently linear [16]. In line with this, neural networks implicated in emotion regulation follow a pattern of protracted refinement and reorganization through synaptic pruning and myelination through late childhood and adolescence into early adulthood [5,11]. This may explicate adolescents' heightened sensitivity to rewarding experiences, increased experience of negative emotions, and variability in affect compared with young children and adults [16-18]. Indeed, strengthening emotion regulation relies on improved connectivity between affective and reward-processing networks and prefrontal cognitive control networks [19]. This key developmental process is malleable and influenced by internal and external factors such as hormonal changes and social relationships [10,11]. Critically, this malleability is somewhat heightened from late childhood through adolescence [5,6].

Emotion Regulation Interventions in Youth

Proximal social input that manipulates the external environment in a positive manner, such as targeted intervention, can improve emotion regulation ability in youth [10]. Traditional face-to-face psychological interventions that aim to promote adaptive emotion regulation in youth include cognitive-, emotion-, and mindfulness-based talking therapies such as cognitive behavior therapy (CBT) [20], rational emotive behavior therapy [21], and dialectical behavior therapy [22]. These are facilitated by a psychologist in 1:1 sessions or small groups, depending upon the needs of the individual and available resources. CBT aims to reduce the selection and implementation of maladaptive cognitive emotion regulation strategies (eg, rumination) and instead promote adaptive ones (eg, cognitive reappraisal). CBT is effective in adolescent populations [23]. However, such therapies are time, money, and personnel intensive [24], and youth may experience traditional programs as unattractive because of perceived mental illness [25,26] and related help-seeking stigma [27]. Negative attitudes toward traditional approaches may be reflected in poor engagement, as evidenced in dropout rates of up to 75% [28].

Preventive emotion regulation programs that are wider reaching than traditional therapies focus on the engagement and education of the caregivers of youth [13]. Such programs encourage explicit tangible learning and practice of adaptive emotion regulation strategies [13] either in the classroom [29-31] or through home-based socialization [32,33]. Although highly encouraging, these interventions may not be accessible or appropriate for all young people. For example, disadvantaged youth demonstrate an increased potential for withdrawal from mainstream services [34] through which wider-reaching interventions are provided.

Digital Interventions in Youth

Mental health digital interventions for youth have attracted a number of recent systematic reviews and meta-analyses [12,35-37]. The most common types of digital interventions include virtual and augmented reality, internet therapy, biofeedback and neurofeedback, digital games, and web-based programs [12,35,36]. Although this is an emerging field, preliminary evidence suggests that digital technologies may constitute clinically effective, economical, accessible, and attractive interventions for mental health problems in youth [12]. Moreover, the internet is widely accessible, even to populations who may not have access to support using traditional means [38].

However, previous reviews have focused on a broad age range, encompassing childhood, early adolescence, late adolescence, and early adulthood. This may not permit the understanding of the impact that digital tools have within childhood and early adolescence—a critical period of brain development [5,6] and time of newly increased interest in, and engagement with, web platforms [39].

This Study

Considering emotion regulation–specific digital interventions, an example is the small number of freely available mobile apps accessible through the UK National Health Service (NHS) digital technology library for mental health. These claim to support well-being through heart rate biofeedback, breathing techniques, and gamified calming strategies. Such freely

XSL•FO
RenderX

available interventions are born out of national health care provision policy implemented by the NHS, which is driven by clinical need and economic considerations; yet, there is no empirical research to provide evidence for the efficacy of these apps. Furthermore, no extant systematic reviews or meta-analyses present such evidence for emotion regulation digital interventions in children and early adolescents.

In parallel to the question of efficacy, the study by Bevan-Jones et al [40] highlighted concerns regarding levels of user engagement, uptake, and adherence in mental health digital interventions for youth. This is discussed in line with best practices in digital intervention development in which active involvement of key stakeholders (eg, early adolescents) is recommended to facilitate the feasibility of digital interventions as well as their acceptability [40]. A systematic understanding of how far digital interventions for emotion regulation have achieved feasibility and acceptability and how these are evaluated is also important. In this systematic review and meta-analysis, we aimed to evaluate the extant evidence base for the use of digital technologies to improve emotion regulation in children and early adolescents and provide recommendations for the progression of the field.

Research Questions

We formulated the following research questions:

1. What are the characteristics of digital interventions that have been evaluated in terms of the efficacy and feasibility of their impact on emotion regulation in children and early adolescents?
2. How efficacious and feasible are emotion regulation digital interventions in children and early adolescents?
3. What are the experiences of children, early adolescents, and other stakeholders regarding the acceptability of emotion regulation digital interventions that evaluate efficacy or feasibility?

Methods

Details of the protocol for this systematic review and meta-analysis were registered on PROSPERO [41].

Information Sources and Search

Web of Science, MEDLINE, PsycINFO, EMBASE, Education Resources Information Centre, ACM Digital Library, and IEEE Xplore electronic databases were used to identify studies. Groups of search terms pertaining to children and early adolescents, digital interventions, and emotion regulation were identified through scoping searches and combined using *OR* (within groups) and *AND* (across groups) Boolean operators and syntax. Search terms and associated Boolean operators and syntax were adapted for different databases as necessary. Refer to Multimedia Appendix 1 for the full search strings for each database. Gray literature searching using Open Science Framework Preprints and OpenGrey electronic databases, as well as forward and backward tracking, was used to identify further studies. An author voluntarily sent 1 study to the authors. Searches were initially run in August 2018 and repeated as a top-up search in July 2020. The initial search was broader than the top-up search. Before the top-up search in July 2020, the inclusion criteria were reviewed. Because of the need to narrow the focus of the review, studies targeting social cognition only and acceptability or qualitative design only, theses, and studies in which samples were aged <8 years or >14 years were excluded. Therefore, at this stage, the social cognition search terms were removed from the search string. The age of 14 years was determined as the upper age limit because of increasing evidence of the need for early emotion regulation intervention efforts from an empirical as well as public health perspective [4,29,30,33,42].

Eligibility Criteria

Studies were included if they met the following inclusion criteria: (1) used digital technology as an intervention strategy, (2) aimed to improve emotion regulation and associated neurobiological mechanisms, (3) targeted children and early adolescents (mean age between 8 and 14 years), and (4) reported data on the efficacy or feasibility of the digital intervention with or without acceptability data. Studies reporting only acceptability data without corresponding efficacy or feasibility data were not included. Studies meeting these inclusion criteria and published after 2008 in peer-reviewed journals presented in English, German, Portuguese, Spanish, Italian, Serbian, Croatian, or Hebrew were considered for inclusion. Studies published after 2008 were included because scoping searches conducted in August 2018 revealed that emotion regulation digital interventions were developed after 2008. Quantitative and mixed methods studies that used any relevant outcome measure were considered for inclusion.

Studies were excluded if they met any of the following exclusion criteria: (1) not original research paper, extended conference paper, or preprint (ie, book, book chapter, commentary, conference abstract, or conference poster), (2) development or testing of technical intervention component only (eg, statistical simulation without assessment of a psychological variable), (3) animal population, and (4) population with organic neurological disorder, and if they did not meet the inclusion criteria.

Please refer to the Meta-analyses subsection under the Methods section for specific information on meta-analysis eligibility criteria.

Study Selection and Data Collection

Using the web-based reference management software Covidence (Veritas Health Innovation Ltd), 2 independent reviewers (SR and JS) conducted record screening [43]. Any conflicts between the reviewers' screening decisions were resolved through consensus, with involvement of a third experienced researcher if necessary. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [44] and a bespoke risk-of-bias assessment tool (refer to the next section) were used when designing and conducting the data extraction protocol. A piloted standardized Microsoft Excel table was used to extract data from the included studies for evidence synthesis and risk-of-bias assessment (refer to Multimedia Appendix 2 [45-83] for the extraction table). The first author (SR) completed this task.

XSL•FO
RenderX

Assessment of Risk of Bias

A tool for risk-of-bias assessment was created using evidence-based information and guidance. This was sourced from the Cochrane Collaboration's tool for assessing risk of bias in randomized trials [84], Cochrane Methods risk-of-bias web-based library [85], NHS National Institute for Health and Care Research guidance for feasibility and pilot intervention studies [86], and the Newcastle-Ottawa Scale for assessing the quality of nonrandomized trials [87]. Wide-ranging information and guidance were required because of the breadth of research designs and associated methodological characteristics included. Care was taken to feature the risk-of-bias domains relevant to the included studies and questions in the review. The tool consisted of 6 domains: (1) selection bias, (2) performance bias, (3) detection bias, (4) attrition bias, (5) reporting bias, and (6) other bias. The risk-of-bias assessment was conducted independently by 2 two trained reviewers. Any conflicts were resolved through discussion, with involvement of a third experienced researcher if necessary. Refer to Multimedia Appendix 3 for the risk-of-bias tool domains.

Clustering and Coding of Included Studies

Guidance on the conduct of narrative synthesis in systematic reviews [88] was followed to generate a thematic understanding of the included digital interventions. The included studies were coded and clustered using defined criteria based on the key category of intervention type. The criteria and definitions that were used to cluster the included studies are detailed in Textbox 1. Studies were further coded based on the population type. These were identified as *diagnosed*: children and early adolescents diagnosed with a physical or mental health disorder; *at risk*: children and early adolescents at risk of a mental health disorder (eg, elevated anxiety); *healthy*: typically developing children and early adolescents with no identified diagnosis; and *universal*: no exclusion criteria applied. The studies' outcome targets were coded based on what they measured. These were *emotion regulation*: recognition of emotions in oneself and the increase or decrease of these emotions; *emotion experience*: negative (eg, frustration) or positive (eg, joy) emotions or symptoms; and *physiological regulation*: brain or bodily signals associated with emotion regulation and emotion experience (eg, heart rate).

Tables were created to summarize the characteristics of the included studies and the efficacy, feasibility, and acceptability data. Within the tables that present the efficacy, feasibility, and acceptability findings, we provide the raw reporting risk-of-bias information at the measure level for information and transparency.

Textbox 1. Clustering of included studies based on intervention type.

Cluster and definition
• Biofeedback: A digital physiological monitoring aid
• Digital game: An electronic game with functions to achieve specific goals, with or without biofeedback
• Virtual reality and augmented reality: A simulated environment (ie, a digital immersion experience with no physical world input); an enhanced reality (ie, a digital sensory component on a live smartphone view)
• Program and multimedia: A program or multimedia application

Acceptability, Feasibility, and Efficacy

Acceptability and feasibility data were synthesized within intervention clusters with validity and reliability reporting bias and attrition bias information where appropriate. In studies not included in the meta-analyses, within-intervention group before-and-after emotion regulation, emotion experience, and physiological regulation efficacy data were further synthesized with significance and effect size information where available. Hedges g (the summary measure) was calculated in R [89], using the esc package [90] to indicate whether significant observed effects were small (0.2), medium (0.5), large (0.8), or very large (1). In a very small number of studies, it was not possible to calculate Hedges g or convert to it (eg, where only η^2 value was provided).

Meta-analyses

Studies included in the systematic review were considered for inclusion in the meta-analytic component if they were randomized controlled trials (RCTs) [91]. Of the 39 studies included in the systematic review, 3 (8%) were not sufficiently homogeneous to other included studies; therefore, they were not included in the meta-analyses. Of these 3 studies, 2 (67%) used informant report only and reported emotion regulation effects alongside emotional expression as a composite score and 1 (33%) implemented a crossover design. In addition, another study did not provide data to calculate effect sizes and hence was not included. Noninferiority RCTs (which compared the intervention to efficacious group face-to-face CBT and hypothesized nonsignificant differences between groups), which are increasingly prevalent in the intervention literature, were included in the meta-analyses. Including noninferiority trials is a more conservative approach in terms of the resultant effect sizes that would be expected; the majority (6/9, 67%) of the other studies included in the meta-analyses used an active control, and some of these were class-based psychoeducation, which has demonstrated beneficial effects. Thus, noninferiority studies were deemed similar enough to other RCTs to be included [92]. It is acknowledged that this may have resulted in the pooled effect being lower than if noninferiority effect estimates had been excluded. Including nonrandomized studies was considered; however, such studies were not reasonably resistant to biases (they were all judged as high risk of bias, inclusive of confounding bias, and varied greatly in methodological design [91]).

Meta-analyses using a very small number of studies may negatively affect the estimation of between-study variance [93];

XSL•FO
RenderX

therefore, a threshold of 4 studies was established as a suitable minimum. Thus, 2 meta-analyses were conducted, focusing on emotion regulation and emotion experience outcomes, respectively. From each study, 1 effect was selected for each meta-analysis to ensure the independence of effect sizes [94]. All the studies used self-report measures; therefore, 1 self-report effect from each included study was selected. Where studies provided multiple self-report effects for each outcome target type, constructs from self-report scales or subscales that were most similar to each other across the included studies were selected. For example, most emotion experience effects measured anxiety across the included studies; hence, where possible, anxiety-based effects were selected for meta-analysis. Where studies used multiple comparison groups, the active control group data were used. Further standardization was facilitated by computing postintervention standardized mean differences only. This is because follow-up data collection was not incorporated into the designs of all meta-analytic studies, and where it was, the length varied greatly across studies. Studies were also coded based on intervention type (biofeedback, digital game, and program and multimedia), population type (diagnosed, at risk, healthy, and universal), training of additional skills (yes or no), use of additional mode of intervention delivery (yes or no), and measure risk of bias (low or high). In addition, dropout rate was included in the results of the meta-analyses.

Analyses were conducted in R [89] using meta-analyses packages tidyverse [95], meta [96], metafor [97], and dmetar [98]. The studies included in the meta-analysis varied somewhat in methodological design; therefore, a degree of heterogeneity was assumed. In line with this, random effects models were applied [99]. The restricted maximum likelihood estimation of tau-squared (τ^2) between-study variance was used because it corrects for negative bias within continuous data (in which large tau-squared is reported when the number of studies and individual studies' sample sizes are small), unlike the standard DerSimonian-Laird method [100]. The mean and SD of each study's selected effect was used to calculate Hedges g and its SEs. Hedges g was computed because the commonly used Cohen *d* [101] may demonstrate a slight bias in small studies in which effects are overestimated [102]. In each meta-analysis, study ID was the unit of analysis, and the effect size (g) for each study was the level of analysis [103].

Heterogeneity was estimated using I^2, tau-squared, and the prediction interval (range into which the effects of future studies are expected to fall) because of the possibility that any one measure on its own is inadequate [98]. Specifically, although I^2 is insensitive to increases or decreases in the number of studies, it relies on each individual study's sample size to predict the amount of variability in the effect sizes *not caused by sampling error* [98,104]. I^2<25% indicates low heterogeneity, I^2=50% indicates moderate heterogeneity, and I^2>75% indicates high heterogeneity [105]. Tau-squared, the between-study effect size variance estimator, is insensitive both to each study's sample size and the number of studies in a meta-analysis, but the meaning of tau-squared might be difficult to interpret alone [98].

Outlier analyses were conducted to determine whether extreme effect sizes contributed to between-study heterogeneity, using a CI-based approach [98]. Influence analyses were conducted to determine the robustness of the pooled effect estimates using leave-one-out principles [98]. Influential cases were examined in subplots [106]. These revealed how much the predicted pooled effect changed in SD units after excluding a given study, the distance between the value when the study was included versus excluded (the Cook distance), and the covariance ratio [98]. Extreme values were shown in red. In addition, the plots were examined to detect any extreme cases not defined by the Viechtbauer and Cheung threshold [106]. Baujat plots were created to determine each study's heterogeneity input [107]. Finally, 2 leave-one-out forest plots that ordered studies by I^2 between-study heterogeneity and effect size (low value to high value) were created to provide further evidence of influential studies [98]. As digital games constituted most (9/11, 82%) of the studies in the meta-analyses, additional meta-analyses were conducted using only digital game effect sizes where appropriate.

Publication Bias

Several steps were taken to investigate potential publication bias, which occurs because of selective publication of significant findings with large effects [98,108]. Particularly in small studies, where very large effects are needed to reach statistical significance, the results are more likely to be statistically significant if their effect sizes are high. First, contour-enhanced funnel plots were examined visually. Contour-enhanced funnel plots, which present color shading linked to significance levels, allow the distinguishing of publication bias from other sources of asymmetry, for example, variable study quality [109]. The Egger test of the intercept quantified funnel plot asymmetry—a statistically significant result (P<.05) determines asymmetry [110]—although this possesses low statistical power in <10 studies [91]. Where the Egger test was significant, the Duval and Tweedie trim-and-fill method was used to estimate the actual effect size had the *missing* small studies been published [111]. Missing studies were imputed into the funnel plot until symmetry was attained.

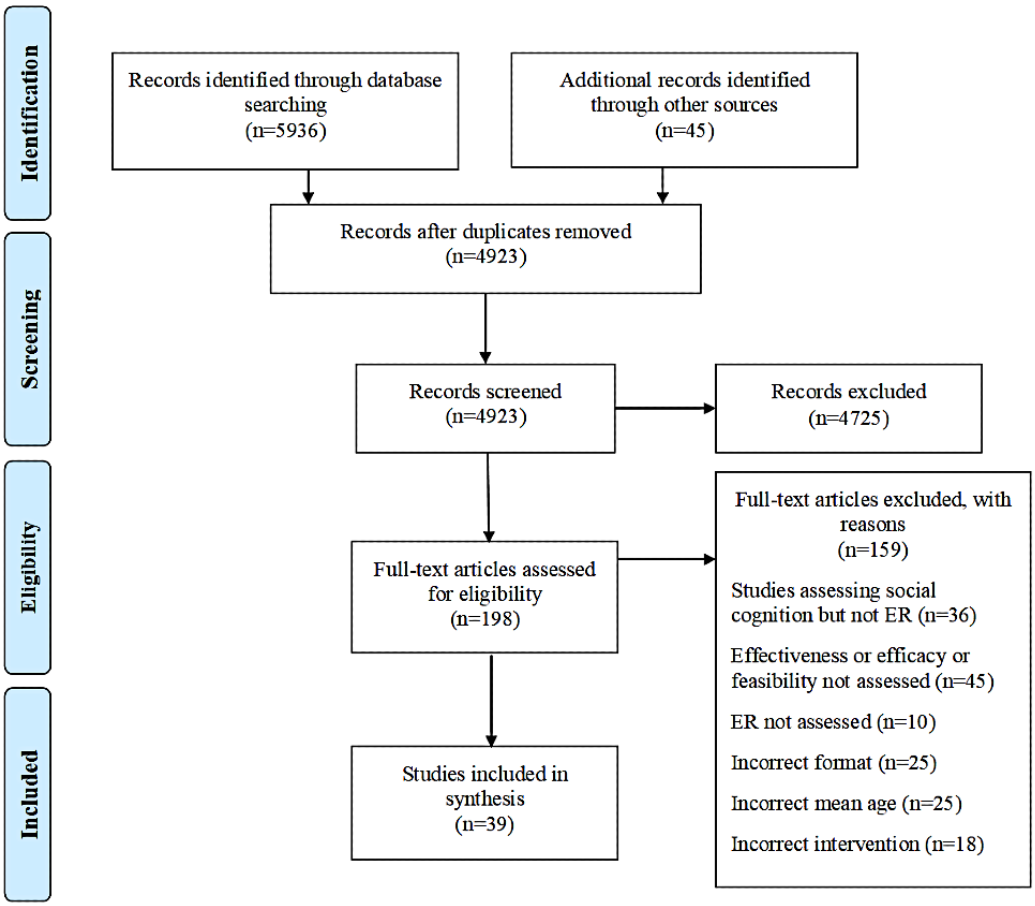
Results

Study Selection

The use of the inclusion and exclusion criteria as previously defined resulted in 39 studies being included in the systematic review and meta-analysis [112] (Figure 1).

XSL•FO
RenderX

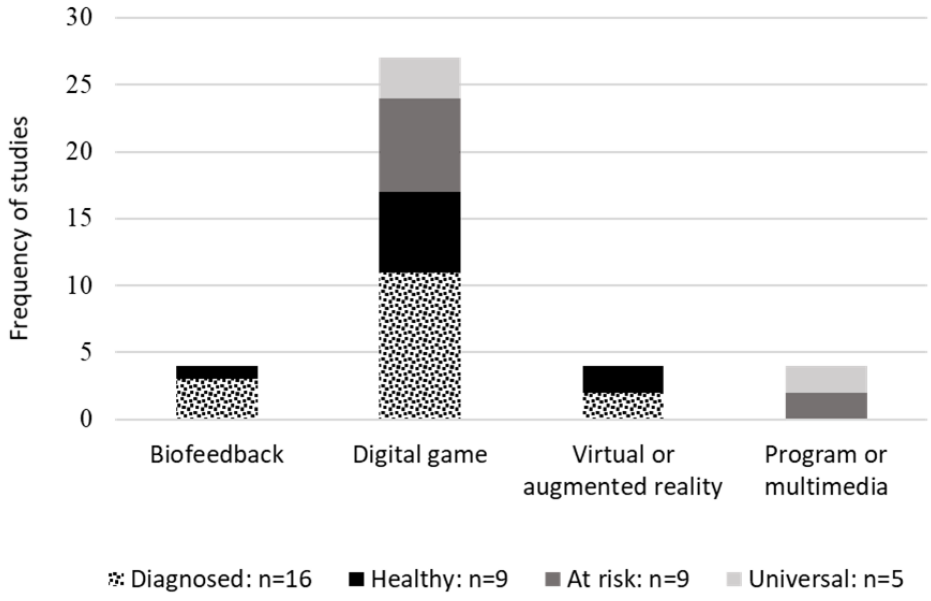
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for study inclusion (adapted from Moher et al [111], which is published under Creative Commons Attribution 4.0 International License). ER: emotion regulation.



Description of Study Clustering

As shown in Figure 2, most (27/39, 69%) of the studies assessed digital game interventions in children and early adolescents who had received a diagnosis.

Figure 2. Study clustering findings with population characteristics. Of the 39 studies, 1 (3%) reported results for both populations who had received a diagnosis and healthy populations. The totals were calculated based on the main target population.



XSL•FO
RenderX

Study Characteristics

Multimedia Appendix 4 [45-83] contains the characteristics of all included studies. The 39 studies had sample sizes ranging from 2 to 1645. Participants were aged 5 to 17 years, with a mean age, where reported, of 8 to 14 years. Studies provided data related to effectiveness (4/39, 10%); effectiveness and feasibility (11/39, 28%); effectiveness and acceptability (2/39, 5%); effectiveness, feasibility, and acceptability (9/39, 25%); efficacy (5/39, 13%); efficacy and feasibility (6/39, 15%); and efficacy, feasibility, and acceptability (2/39, 5%). Of the 17 studies that targeted children and early adolescents who had received a diagnosis, most (n=9, 53%) targeted autism spectrum disorders (ASDs). Of the 8 studies that targeted samples classified as *at risk*, half (n=4, 50%) targeted elevated anxiety with digital games. Studies were conducted in Australia (9/39, 23%), Spain (8/39, 21%), The Netherlands (6/39, 15%), the United States (6/39, 15%), Hong Kong (3/39, 8%), Romania (3/39, 8%), Wales (1/39, 3%), Nepal (1/39, 3%), Belgium (1/39, 3%), and Germany (1/39, 3%). Differentiation between effectiveness and efficacy highly depends on study design and available resources; indeed, effectiveness reflects real-life conditions. Hence, throughout the reporting of the results, we use the term *efficacy* for simplicity.

In total, 11 studies were eligible for meta-analyses. These comprised 2476 participants (n=1248, 50.4%, in intervention conditions and n=1228, 49.6%, in control conditions). Sample sizes ranged from 20 to 1645. Most of the studies targeted samples classified as at risk or diagnosed (8/11, 73%) and were digital games (9/11, 82%). Of the 9 digital games, 4 (44%) targeted children or early adolescents at risk of anxiety (n=448); 3 (33%) targeted those diagnosed with posttraumatic stress disorder, anxiety with and without comorbid intellectual disability, and ASD with elevated anxiety (n=178); and 2 (22%) targeted healthy early adolescents (n=185). Of the 11 studies, 1 (9%) biofeedback study targeted youth (n=20) diagnosed with anorexia nervosa, whereas 1 (9%) program and multimedia study targeted a universal sample (n=1645). No studies in the virtual and augmented reality cluster were included. Regarding comparisons, the digital game cluster compared the intervention with an active control (n=2), active control with (n=1) and without (n=2) treatment as usual, active control with separate wait-list (n=2), and treatment as usual with (n=1) and without (n=1) wait-list. The biofeedback study compared the intervention with treatment as usual. The program and multimedia study compared the intervention with a web-based neuroscience program. Of the 5 studies that permitted continuance of usual treatment, 3 (60%) were in the digital game cluster.

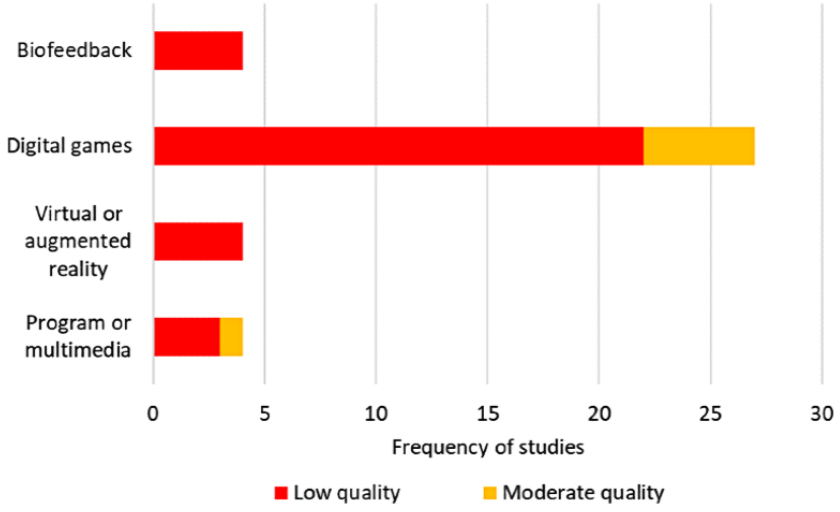
Intervention Characteristic Summary

Of the 39 included studies, 22 (56%) clearly stated that they incorporated additional support, monitoring, or nondigital delivery. Most (3/4, 75%) of the program and multimedia studies incorporated class sessions and homework. Half (4/8, 50%) of the virtual and augmented reality studies as well as biofeedback studies and 56% (15/27) of the digital game studies incorporated nondigital delivery, additional support, or monitoring. Only the study by Wijnhoven et al [45] was included in the meta-analytic component. In total, 44% (17/39) of the studies trained other skills as well as emotion regulation. These were mostly (9/17, 53%) social skills and social cognition. A key pedagogical and therapeutic theme across all interventions was explicit emotion regulation strategy learning through digital characters or a face-to-face facilitator, with practice in a relevant and engaging but safe environment. Refer to Multimedia Appendix 5 for descriptions of all the interventions.

Risk of Bias

As demonstrated in Figure 3, most (33/39, 85%) of the studies were judged by reviewers as low quality. In total, 15% (6/39) of the studies, all of which were included in the meta-analytic component, gained moderate quality ratings. Although most (17/39, 44%) of the studies targeted diagnosed populations, reviewers judged all these studies as low quality. The highest proportion of moderate quality ratings was in the digital game cluster in populations classified as at risk. Overall, the distribution of risk-of-bias scores ranged from 7 to 20 out of 26, with higher scores indicating higher quality. The interrater reliability was substantial (Cohen κ=0.75) [113]. Refer to Multimedia Appendix 6 for a detailed summary of the risk-of-bias findings.

Figure 3. Review authors' judgments regarding overall study quality in the intervention clusters.



XSL•FO
RenderX

Meta-analysis

Emotion Experience

Of the 39 included studies, 10 (26%) assessed group differences in emotion experience with self-report. Of these 10 studies, 9 (90%) revealed effect sizes in favor of the intervention, with less negative (*k*=8) or more positive (*k*=1) emotion experience effects in the intervention group. However, of these 10 studies, only 1 (10%; digital game) revealed a significant effect (Table 1). This study targeted children at risk of anxiety. Only the study by Lackner et al [68] revealed an effect in the unexpected direction in which negative emotion experience was greater in the intervention group than in the control group after the intervention. This related to the only biofeedback study included in the meta-analysis, with the smallest sample size (n=20; although the study also reported one of the lowest dropout rates of 9%). The very small pooled effect was nonsignificant (*k*=10; Hedges g=–0.12, 95% CI –0.26 to 0.02; *P*=.09; Figure 4A). Tau-squared was low (τ^2=0.0176) indicating little variation among the studies. Yet, the I^2 value of 39.5% indicated low to moderate heterogeneity, and the somewhat broad prediction interval (–0.46 to 0.22) suggests that the very small observed pooled effect largely on negative emotion experience through emotion regulation digital interventions is not robust in every context.

Given the potential impact of the type of digital intervention on the pooled effect and heterogeneity outcomes, the emotion experience meta-analysis was conducted again with only the digital games studies (n=9). All the digital game studies assessed negative emotion experience outcomes. The forest plot reveals a small negative pooled effect (Figure 4B). This was significant (*k*=8; Hedges g=–0.19, 95% CI –0.34 to –0.04; *P*=.02). Tau-squared was 0, indicating that variation in effect sizes among the studies was caused by sampling error rather than heterogeneity. The I^2 value of 0% corroborated this, and the narrow prediction interval (Hedges g=–0.34 to –0.04) suggests that the small observed pooled negative emotion experience effect through emotion regulation digital game interventions is robust across different contexts.

XSL•FO
RenderX

Table 1. Emotion experience and emotion regulation meta-analytic outcomes. Refer to Multimedia Appendix 7 [45-83] for measure details[a].

Study, year [reference number]; intervention	Hedges g (95% CI)		Sample size, n	Control	Measures and risk of bias		Dropout rate, %	Other skill or support
	Emotion regulation	Emotion experience			Emotion regulation	Emotion experience		
Lackner et al[b], 2016 [68]; biofeedback	0.41 (–0.48 to 1.3)	0.07 (–0.81 to 0.94)	20	TAU[c]	ECQ R[d,e]	BSI A[f,e]	9	No
Scholten et al[g], 2016 [73]; digital game	—[h]	–0.11 (–0.45 to 0.22)	138	Active	—	SCAS-C[i,e]	8.7 (ITT[j])	No
Schuurmans et al[b], 2018 [75]; digital game	—	–0.13 (–0.78 to 0.51)	37	TAU	—	SCAS-C[e]	34 (ITT)	No
David et al, 2019 [58]; digital game	0.37 (–0.03 to 0.77)	–0.28 (–0.68 to 0.12)	96	Active	ERICA C[k,e]	SDQ-C E[l,e]	7	No
Rogel et al[b], 2020 [71]; digital game	—	–0.70 (–1.41 to 0.02)	32	TAU-WL[m]	—	TSC A[n,e]	22 (ITT)	Executive function
Schoneveld et al[g], 2016 [46]; digital game	—	–0.42 (–0.76 to –0.08)[o]	136	Active	—	SCAS-C[e]	25.7 (ITT)	No
Schoneveld et al[g,p], 2018 [47]; digital game	—	–0.03 (–0.32 to 0.27)	174	Active	—	SCAS-C[e]	12 (ITT)	No
David et al, 2020 [59]; digital game	—	–0.10 (–0.51 to 0.32)	89	Active	—	PoAD A[q,r]	18.8	No
Wijnhoven et al[b], 2020 [45]; digital game	—	–0.14 (–0.51 to 0.24)	109	Active	—	SCAS-C[r]	32	Therapist
Schoneveld et al[g,p], 2020 [74]; digital game	–0.11 (–0.4 to 0.19)	—	174	Active	SEQ SE[s,e]	—	12 (ITT)	No
Smith et al[b], 2018 [77]; program	0.26 (0.16 to 0.36)[t]	0.07 (–0.02 to 0.17)	1645	Active	ATES[u,r]	EWS[v,r]	0 reported	Adaptive attitude toward emotion regulation

[a]Pooled Hedges g (random effects model, restricted maximum likelihood tau-squared): emotion regulation: Hedges g=0.19 (95% CI –0.16 to 0.54); emotion experience: Hedges g=–0.12 (95% CI –0.26 to 0.02), game only Hedges g=–0.19 (95% CI –0.34 to –0.04).

[b]Continuance of existing treatment permitted.

[c]TAU: Treatment as usual.

[d]ECQR: Emotional Competence Questionnaire, Regulating and Controlling Own Emotions subscale.

[e]Low risk of bias.

[f]BSIA: Brief Symptom Inventory, shortened from Symptom Checklist-90-Revised, Anxiety subscale.

[g]Continuance of existing treatment not permitted.

[h]Not available.

[i]SCAS-C: Spence Children's Anxiety Scale.

[j]ITT: intention to treat used.

[k]ERICA C: Emotion Regulation Index for Children and Adolescents, Control subscale.

[l]SDQ-CE: Strengths and Difficulties Questionnaire–Child Version, Emotional Symptoms subscale.

[m]TAU-WL: Treatment as usual–waitlist.

[n]TSCA: Trauma Symptom Checklist for Young Children, Anxiety scale.

[o]Significant at $P<.05$.

[p]Noninferiority: no significant between-group differences expected.

[q]POAD A: Profile of Affective Distress, Concern and Anxiety subscale.

[r]High risk of bias.

[s]SEQ SE: Self-Efficacy Questionnaire for Children, Emotion Self-Efficacy scale.

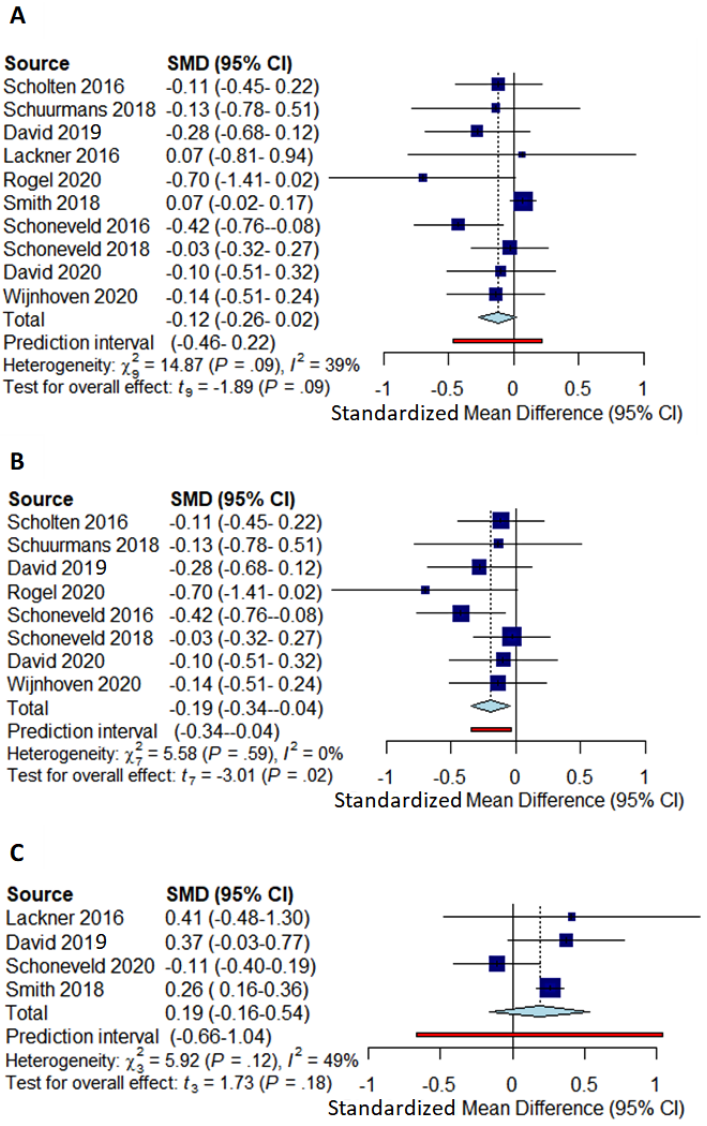
[t]Significant at $P<.01$.

[u]ATES: Adaptive Theories of Emotions Scale.

[v]EWS: Emotional Well-Being in School Scale.

XSL•FO
RenderX

Figure 4. Meta-analytic forest plots (random effects model, Hedges *g*, restricted maximum likelihood tau-squared): (A) Emotion experience. (B) Emotion experience—digital game studies only. (C) Emotion regulation.



Emotion Regulation

Of the 11 studies included in the meta-analysis, 4 (36%) assessed group differences in emotion regulation with self-report. Only the study by Smith et al [77] (program) revealed a significant effect (Table 1). The non-inferiority study by Schoneveld et al [74] revealed an effect in favor of the control group. Of note is the biofeedback study by Lackner et al [68] in which the control group improved compared with the intervention group; yet, because the intervention group's baseline mean was greater than that of the control group, the observed effect seems to be in favor of the intervention group. The pooled effect was nonsignificant (k=4; Hedges g=0.19, 95% CI –0.16 to 0.54; P=.18; Figure 4C). Tau-squared was low (τ^2=0.0274), suggesting little variation among the studies. However, the I^2 value of 49.3% indicated near-moderate heterogeneity, and the extremely broad prediction interval (–0.66 to 1.04) suggests that the nonsignificant small observed pooled effect on emotion regulation through emotion regulation digital interventions is not robust.

Outliers and Influential Cases

Outlier analysis did not detect any extreme effect sizes for the emotion experience or emotion regulation meta-analyses.

In the meta-analysis on emotion experience outcomes in digital games (significant), no studies were identified as extreme cases using the influential Viechtbauer and Cheung study threshold [106]; yet, visual inspection of the influence analysis subplots suggested that the studies by Schoneveld et al [46] and Schoneveld et al [47], both of which trained emotion regulation with an electroencephalogram (EEG) neurofeedback–based anxiety-induction digital game, presented extreme values. The Baujat plot corroborated this, indicating that these studies were highly influential in heterogeneity and pooled effect size. These studies also measured efficacy expectancy before the intervention and reported null between-group differences. Refer to Multimedia Appendix 8 [46,74,77,106] for a detailed description of the influence analyses.

In summary, the meta-analytic evidence suggests that only digital game interventions significantly reduced negative emotional experience in children and early adolescents with a

XSL•FO
RenderX

small effect, and this may be robust across different contexts; yet, there is no evidence for improvements in self-reported emotion regulation abilities through digital intervention.

Publication Bias

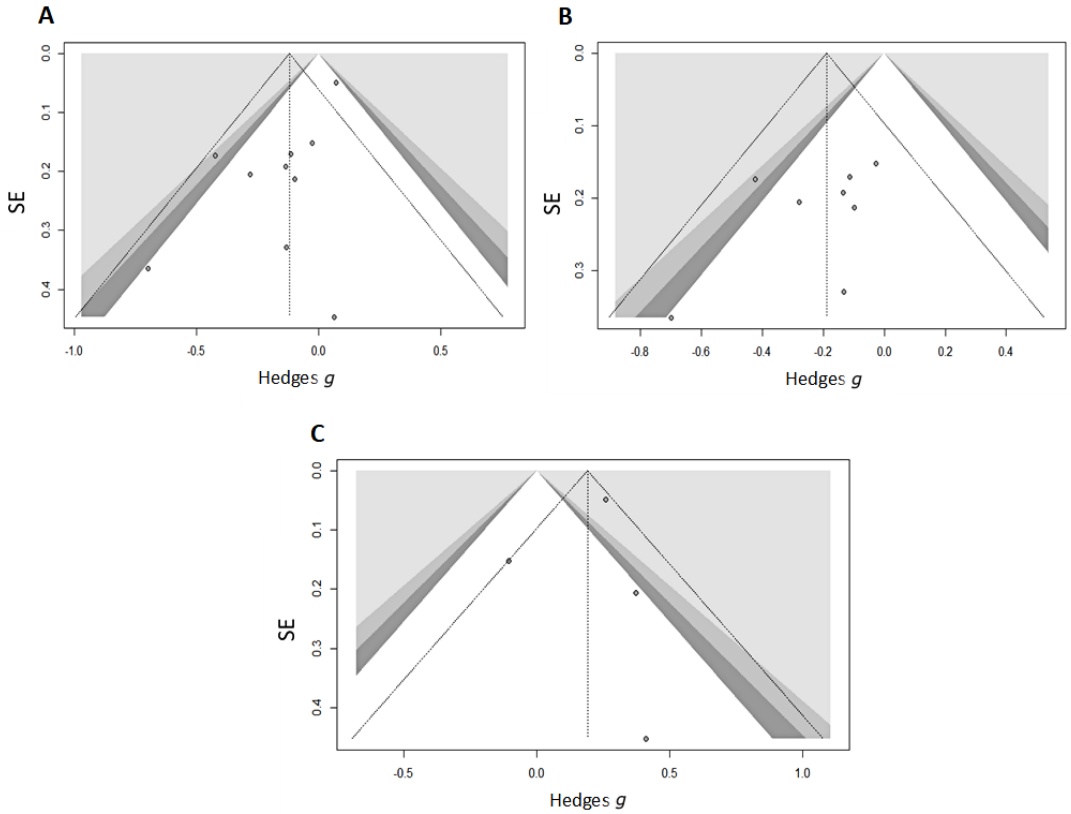
Visual inspection of the contour-enhanced funnel plots (Figure 5) indicated some asymmetry. Importantly, there was only 1 significant effect size in each funnel plot. This suggests that asymmetry may have been due largely to factors other than publication bias (eg, variations in study quality and methodology).

The Egger test of the intercept was nonsignificant for both the emotion regulation (*k*=4; regression intercept –0.409, 95% CI –3 to 2.19; *P*=.79) and emotion experience digital game–only meta-analyses (*k*=8; regression intercept –1.514, 95% CI –3.87 to 0.84; *P*=0.25). However, it was significant for the emotion experience meta-analyses that included all relevant studies (*k*=10; regression intercept –1.462, 95% CI –2.36 to –0.57; *P*=.01). Hence, there was substantial asymmetry within this funnel plot potentially because of variations in study quality and methodology.

A trim-and-fill analysis was conducted on the significant emotion experience effect (refer to Multimedia Appendix 9 for the associated funnel plot). The 5 added effects were larger in magnitude, and the pooled effect was smaller and remained nonsignificant (*k*=15; Hedges *g*=–0.019, 95% CI –0.16 to 0.2; *P*=.82). Tau-squared was moderate (τ^2=0.05), indicating variation among the studies. I^2 was 56%, indicating moderate heterogeneity.

In summary, the small, significant impact of *digital games* on negative emotional experience in children and early adolescents was likely *not* overestimated because of either publication bias or variations in study quality and methodology.

Figure 5. Meta-analytic contour-enhanced funnel plots between the SE and Hedges *g*. (A) Emotion experience. (B) Emotion experience—digital game studies only. (C) Emotion regulation. Light gray shading: *P*<.01; gray shading: *P*<.025; and dark gray shading: *P*<.05. No shading: nonsignificant (*P*<.05).



Efficacy

Overview

Multimedia Appendix 10 [48-57,60-67,69,70,72,76,78-83] contains the within-intervention group pre- to postintervention efficacy summaries of emotion regulation, emotion experience, and physiological regulation domains from all studies not included in the meta-analyses (28/39, 72%). Where available, follow-up data are also provided. Studies were nonrandomized or noncontrolled or judged to not be adequately resistant to biases and of variable methodological design. Therefore, the synthesis assessments described in the following sections should be interpreted accordingly. Where it was not possible to synthesize before-and-after efficacy data (eg, single-session experiments, postintervention interviews, and field notes), data were synthesized in line with the measures and design from which they were borne.

Biofeedback Studies

Of the 3 biofeedback studies that provided efficacy data, only 1 (33%) used measures that were judged as low risk of bias. All 3 studies provided data on physiological regulation. Children and early adolescents significantly altered their physiology as directed by the intervention in heart rate variability (HRV)–EEG

XSL•FO
RenderX

and functional magnetic resonance imaging (fMRI) biofeedback-neurofeedback and neurofeedback interventions. Emotion regulation was assessed in 67% (2/3) of these studies: emotion regulation correlated with increased emotion regulation network activation in the fMRI neurofeedback study, whereas emotion regulation improved significantly with a large effect in HRV biofeedback but not in combined HRV-EEG biofeedback-neurofeedback in the second study. This study also assessed negative emotion experience—emotional lability and negativity decreased significantly with a large effect. Anxiety reductions were nonsignificant.

Digital Game Studies

Of the 27 digital game studies, 18 (67%) provided efficacy data. Of these 18 studies, 3 (17%) assessed the success of frustration or joy emotion induction within a virtual reality–enabled emotion regulation game across different mediating devices. Frustration increased significantly after the frustrating game in 33% (1/3) of the studies but not when mediated by a camera device. Increases in joy after the joyful game were nonsignificant across all device types in 33% (1/3) of the studies.

The strongest evidence for positive change brought about by digital games was the reduction of negative emotion experience (anxiety). Of the 18 studies, 7 (39%) measured this; of these 7 studies, 3 (43%) were statistically significant with small to large effects.

Of the 18 studies, 8 (44%) measured emotion regulation, which largely improved. Where pre-post statistical information was available (5/8, 63%), improvements were significant, with medium to large effects. However, the significant findings reported on the same game.

Of the 18 studies, 5 (28%) assessed physiological regulation. In total, 20% (1/5) of the studies reported significant reductions in heart rate and 20% (1/5) reported nonsignificant reductions in heart rate.

Virtual and Augmented Reality

Of the 4 virtual and augmented reality studies, 4 (100%) provided efficacy data, largely with measures judged as low risk of bias. Most (3/4, 75%) of the studies only measured emotion regulation. Individual exposure but not group exposure to immersive virtual reality emotion and social skill practice was linked to significant improvements in emotion regulation in a sample with high-functioning ASD, with a small effect (from 2/4, 50%, studies).

Program and Multimedia

Of the 4 program and multimedia studies, 3 (75%) provided efficacy data, all of which assessed 1 multimedia modular program. Intensity of emotions was assessed in 67% (2/3) of these studies—intensity of negative emotions only decreased significantly in 50% (1/2) of these studies, with a small effect. Intensity of positive emotions decreased significantly in both studies, with small to large effects.

Summary of Efficacy Data

The most consistent evidence comes from digital game interventions in the reduction of negative emotion experience. A note of caution is recommended when interpreting these findings owing to the varied methodology, high risk of bias, and overall low quality of the included studies. Furthermore, the evidence base for the impact of digital interventions on physiological indices of emotion regulation is much smaller and less consistent.

Feasibility

Multimedia Appendix 11 [45-51, 53, 54, 56, 58, 59, 65, 66, 68-79, 82, 83] contains the feasibility summaries from the included studies. Feasibility data were provided for 72% (28/39) of the studies. All studies that provided feasibility data used measures judged as high risk of reporting bias.

Of the 28 studies, 18 (64%) digital game studies provided feasibility data for various aspects of feasibility, including engagement, implementation, adherence, expectations, and transference to real life. Where the dropout rate was particularly high (>30%), studies targeted samples who had received a diagnosis and the dropouts were largely attributed to personal or family issues. Most feasibility issues were in early-stage small studies (3/18, 17%) in which interventions were prototypes not previously evaluated or were delivered by individuals inexperienced in the intervention technology.

All (4/4, 100%) the virtual and augmented reality studies provided feasibility data encompassing engagement, implementation, and transference to real life. Of these 4 studies, 2 (50%) reported dropout rates, and these were very low.

Of the 4 program and multimedia studies, 3 (75%) provided feasibility data encompassing implementation and engagement. Of these 3 studies, 2 (67%) reported dropout rates, and these were very low.

In summary, most feasibility issues were in early-stage interventions targeting samples who had received a diagnosis. Digital game interventions that incorporated biofeedback provided the most evidence for transference of learned emotion regulation skills to real life. However, digital games also presented the highest dropout rate, and all measures across all clusters were judged as high risk of reporting bias.

Acceptability

Multimedia Appendix 12 [46-49,53,54,65,67,69,72,75,80,82] contains the acceptability summaries from the included studies. Acceptability data were provided for 33% (13/39) of the studies. The biofeedback cluster did not contain acceptability data.

Of the 9 digital game studies that measured acceptability, 6 (67%) reported moderate to highly positive results for at least one aspect of acceptability, including likability, flow, usability, helpfulness, difficulty, appeal, usefulness, and relevance. The only study that reported mainly negative acceptability findings highlighted a link between guided imagery, visualization, and deep breathing games being too difficult or easy and poor likability in children diagnosed with ASD.

Of the 4 virtual and augmented reality studies, 2 (50%) that evaluated acceptability in 2 interventions reported mainly positive findings for fun, educational impact, likability, motivation impact, and experienced happiness. The only study

XSL•FO
RenderX

across all clusters that used a measure judged as low risk of reporting bias assessed an outdoor augmented reality quest (which involved meeting other players). Importantly, it was viewed as potentially dangerous, although the authors did not elucidate exactly to what this *danger* pertained.

Of the 4 program and multimedia studies, 2 (50%) that evaluated acceptability in a school-based program reported high likability and a moderate educational impact.

In summary, emotion regulation digital interventions were largely acceptable to children and early adolescents, as well as other key stakeholders. However, of the 20 measures, 19 (95%) were judged as high risk of reporting bias. Negative acceptability findings were mainly in small early-stage digital game interventions targeting samples who had received a diagnosis.

Discussion

Summary

This systematic review and meta-analysis aimed to evaluate current digital interventions that train emotion regulation in children and early adolescents published in peer-reviewed articles up to July 2020. In summary, digital games were the most prevalent intervention type: 69% (27/39) of the studies evaluated digital games. Digital games decreased negative emotional experience with a small significant effect, mainly in samples at risk of anxiety. In addition, digital interventions improved emotion regulation; yet, this effect was nonsignificant. Furthermore, acceptability was strong across all intervention types and samples, and most feasibility-related problems were in samples who had received a diagnosis. In the following sections, we discuss the key findings and provide recommendations for the field's progression.

Efficacy

Examined through meta-analysis and systematic review, digital games provided evidence for a significant reduction in negative emotional experience with a small effect, largely in samples at risk of anxiety, using validated and reliable outcome measures. This suggests that digital games are the most advanced and efficacious digital interventions for training emotion regulation in children and early adolescents. This important finding may be partly explained with cognitive load theory, which postulates that limited novel information can be processed at once in working memory [114]. Indeed, to optimize learning in a digital environment, balance must be sought between presenting information in a manner that meets an individual's cognitive needs, yet with sufficient complex information to facilitate understanding of the given topic, and learning must be active to enhance the development of cognitive schemas [114,115]. Such optimization may be achieved with certain pedagogical techniques. For example, pacing serves to decrease cognitive load on working memory by relying on the user or system to control information presentation (eg, by pausing material delivery or going back to look at previous material) [115]. In line with these digital pedagogical principles, the included digital game studies largely presented learning tasks that focused on different emotion regulation strategies within separate parts of the game, with gradual user-led increases in difficulty and complexity, and a simple user-friendly interface, with animated characters that provided information about different emotion regulation strategy elements and in-game support.

In combination with digital game design methods that optimize cognitive flow [116], feelings of autonomy [117,118], and fun [119], digital game training may have increased motivation and engagement, which are recognized barriers to efficacy in digital interventions in children and early adolescents [40].

Neurofeedback may also be key to this finding; meta-analytic influence analysis indicated that the digital game studies that incorporated EEG neurofeedback (2/27, 7%) drove the small significant pooled effect. This is in line with the embodied emotion regulation framework [120], which proposes a distinction between cognitively based top-down (cognitive labeling, mindful detachment, meta-awareness, and cognitive reprisal) and affect-driven bottom-up (sensory perception and interceptive proprioception) emotion regulation strategies and argues that they work together as part of an integrated emotion regulation system. Hence, it is possible that these interventions successfully addressed both top-down and bottom-up strategies, which increased efficacy. In addition, the real-time visual neurofeedback may have further increased immersion within the digital game and, subsequently, engagement [40]. However, neurofeedback information provided to players was collected using non–research-grade EEG equipment, and double blinding was not incorporated. In this context, the role of placebo effects on the apparent impact of neurofeedback on clinical symptomatology must be considered. This was discussed by 3 studies [121-123] in line with prior clinical neurofeedback research in which diligent methodological rigor is not evident; yet, significant intervention effects are routinely reported. The emotion regulation digital intervention field should address concerns around potential placebo effects in neurofeedback through the application of methodological rigor, including double-blind, placebo-controlled trials [122].

When planning placebo-controlled trials it is important to consider that expectations around intervention effects may influence placebo effects; yet, such expectations are rarely measured in light of this [124]. The higher-quality digital game studies (2/27, 7%) in this review that drove the emotion experience findings measured intervention expectation at baseline and reported null between-group effects. However, earlier-stage studies, not included in the meta-analytic component, did not. As the field progresses, intrinsic motivation must be harnessed in double-blind, placebo-controlled trials, with expectancy measured at baseline, particularly in studies that incorporate neurofeedback components. In addition, although portability and ease of use drive the use of non–research-grade EEG equipment, it is argued that such issues must be balanced against the impact on the credibility of the tool. That is, if digital emotion regulation training in children and early adolescents relies on suboptimal technology, are we really driving the field forward?

There was limited measurement of emotion regulation across all included studies; hence, the available emotion regulation efficacy findings must be interpreted with caution. The lack of focus on emotion regulation may be due to the included studies

XSL•FO
RenderX

focusing somewhat on children and early adolescents at risk of *anxiety* and the concurrent training of *social cognition* and *social skill* difficulties; hence, these constructs were the key outcomes. In addition, there are limited psychometrically sound emotion regulation measures for children and adolescents, despite increasing awareness of the importance of its adaptive development [125]. To advance the field, there is a requirement for researchers to create and validate emotion regulation measures for diverse child and early adolescent samples, and digital intervention studies should objectively assess improvements in emotion regulation ability after the intervention and at follow-up.

Considering emotion regulation *knowledge*, medium to large significant improvements were observed in digital game studies that also applied additional therapeutic support, parental guidance, or targeted social cognitive skills, particularly in samples with ASD. Indeed, research has highlighted associations between brain regions implicated in cognitive emotion regulation and social cognition in youth [126] and the requirement of perspective taking [127] and abundant semantic representations [128] for successful alternative representations of emotion-inducing stimuli (ie, cognitive reappraisal). Furthermore, the integration of caregivers in interventions for samples with ASD may boost the generalizability of learned skills [129] and increase engagement with the intervention [130]. Hence, the inclusion of social cognition training as well as caregiver support may have positively influenced the emotion regulation improvements observed in these studies. Therefore, it may be beneficial to include social cognitive training and parental support within emotion regulation digital interventions that target samples with ASD because this may enhance their efficacy. However, as emotion regulation knowledge improvement was only assessed in a small number of lower-quality studies, we recommend that caution must be taken when interpreting such findings.

Feasibility

Small early-stage digital game studies that targeted ASD, attention-deficit/hyperactivity disorder, and samples with undefined emotional disorders identified several important feasibility issues linked to generalization, implementation, technical issues, and physiological and emotional symptoms. Interindividual variability and related intervention difficulties are common in samples with neurodevelopmental disorders [131]. Hence, greater feasibility difficulties in such samples are expected. Furthermore, had feasibility issues not been picked up at this early stage of evaluation, full-scale evaluation may have yielded less favorable findings. A program and multimedia intervention assessed in slightly larger studies (2/39, 5%) found that the content was too complex for children at risk of exclusion or suspension from school, and postintervention reductions in the intensity of emotions were variable. Had the ability of the target sample to understand intervention content been checked at an early intervention development stage, efficacy outcomes might have been more consistently positive because the key messages would have been better understood. The relative importance of early-stage studies is emphasized through the consideration of the Medical Research Council's guidelines for complex intervention development [132]. Here, the impact of contextual factors on intervention success is highlighted and has recently been discussed further in a digital intervention context [40]—it is advised that iterative feasibility assessments that examine the issues revealed throughout intervention development are key to understanding contextual factors.

A further key finding was the higher dropout rate in samples who had received a diagnosis, especially in the digital game cluster. It is possible that because digital games made up a high proportion of the included studies, they also presented the most realistic picture of dropouts in digital interventions for emotion regulation. Moreover, because digital games were largely evaluated in terms of their effectiveness in real-world settings (eg, school, home, and inpatient care) this may have affected adherence and, subsequently, dropouts. Certainly, adherence to digital interventions outside of research settings in children and early adolescents is an extant key issue [12,133]. Involving the population in the design process who will ultimately use the digital intervention may lead to the iterative development of tools that are feasible and address the high dropout rate [40]. Methods to involve youth in digital intervention development may be optimized to increase engagement [40]. These include using progress bars, animations, and multiple platforms in web-based questionnaires; usability *think-aloud* protocols instead of standard interviews (observed and/or interviewed simultaneously while using the intervention); clear rules and use of materials (eg, screens and devices) in focus groups; and principles applied to focus groups with *wall storms* (sticky notes on a wall) and *word clouds* (grouping of key words) in participative workshops [40].

As digital games that incorporated biofeedback provided the greatest evidence for generalizability of learned emotion regulation skills, this suggests that biofeedback-based digital games may be the most appropriate emotion regulation digital intervention for transference to real life; yet, there is no extant empirical research to support this. Objective assessment of generalization was only conducted in biofeedback-based digital games in samples who had received a diagnosis *(psychiatric and neurodevelopmental disorders)*; hence, this finding may simply be an artifact of the relative prominence of biofeedback-based digital game interventions and inadequate measurement of generalization in the other included studies, although it is important to consider the significance of the specific emotion regulation strategies—deep breathing and cognitive emotion regulation—that seemed to demonstrate the greatest real-life generalizability in children and early adolescents who had received a diagnosis. fMRI-based and self-report–based evidence in adult populations suggests that cognitive reappraisal may be linked to future rather than immediate emotion regulation success in reducing negative emotion (ie, when emotion-inducing stimuli are re-encountered at a later date) [134]. This suggests that the real-life relevance of content within digital cognitive emotion regulation training may be particularly important such that it should clearly relate to the target samples' *real-life experiences and difficulties* to promote future use of learned strategies. In addition, higher cognitive reappraisal frequency is linked to reduced risk for psychiatric symptomatology [135], optimal academic attainment

XSL•FO
RenderX

[136], social outcomes [137], and psychological well-being [138].

Researchers should collaborate with key stakeholders to create highly relevant and engaging intervention components of appropriate complexity to produce improvement in indices of emotion regulation with real-world generalizability [40,139]. Importantly, generalizability should be consistently assessed to determine the emotion regulation digital interventions that are most appropriate across different child and early adolescent samples.

Acceptability

The only study that provided solely negative acceptability data assessed emotion regulation mini-games in an early-stage small evaluation. Here, poor likability was linked to unsuitable difficulty for individuals' needs in a sample with high-functioning ASD. As mentioned previously, a key factor in the presentation of ASD and associated interventions is interindividual variability and related intervention difficulties [131]. Digital technologies allow for greater person-centered training through the involvement of caregivers and the ability to engage with the intervention at home [131]. However, if the caregiver is not able to quickly and easily adjust the difficulty of the intervention or if it is not programmed to adapt dynamically, as also reported in the highlighted study, such caregiver involvement may be in vain [140]. Hence, it is recommended that emotion regulation digital games, especially those designed for children and early adolescents with neurodevelopmental disorders, should incorporate game mechanics that adapt dynamically to an individual's needs, permitting increases and decreases in difficulty as required in real time. For this to be successful, interdisciplinary collaboration is required at all stages of conceptualization, specification, and programming [141]. Specific collaborators may include psychologists, cognitive neuroscientists, educators, therapists, engineers, and, principally, the target of the intervention (ie, youth) [40,139,142].

A further notable finding was the importance of relevance in emotion regulation digital games. Specifically, in an immersive EEG-neurofeedback and anxiety-induction game set within a haunted mansion with ghosts, the experience of relevance to real life was significantly less than that in a group-based CBT comparator but not less than that in a nontherapeutic commercial game comparator. Hence, it may be important to include explicit training content in emotion regulation games that clearly relates to the target samples' *real-life experiences and difficulties* and encourages children and early adolescents to practice learned skills in their daily lives, optimize acceptability, and encourage generalization, as practiced in traditional talking therapies (eg, CBT and dialectical behavior therapy) [20,22]. However, appeal and flow were also key to the experience of acceptability in emotion regulation digital games—the evidence suggests that the experience of these aspects of acceptability may be inferior in emotion regulation games compared with commercial games. Consequently, because relevance, appeal and flow may come into conflict in emotion regulation digital game acceptability, it is recommended that a balance between them should be struck to optimize acceptability. This requires iterative codevelopment at all stages of evaluation [40,139].

Of vital importance to any research activity is the safety of participants, both objectively and through their own subjective experience. Perhaps reflective of the limited acceptability evaluation yet great variability in the type of acceptability assessed in the included studies, only 3% (1/39) of the studies assessed feelings of safety—a commercial augmented reality outdoor-based quest (Pokémon GO) found that participants experienced high levels of perceived danger when engaging in the intervention. Although details were not reported, it is sensible to construe that this may be in relation to the potential for harm from strangers because of interaction with unknown players. Hollis et al [143] provided an overview of the extant cultural and political debate and related research concerning the role and impact of digital technology in the lives of youth. Describing it as a *triple-edged sword*, the authors stated that it fosters personal development and growth; may detect and address mental health issues; and yet could pose purported social, intellectual, and mental health risks. This debate is increasingly heightened because digital technology (and the means to access it) is more important than ever in supporting the educational and socioemotional needs of youth through the COVID-19 pandemic. Crucially, most social, intellectual, and mental health concerns around the impact of new digital technologies—largely driven by population as well as political and academic arenas—may be challenged through nuanced research examining the impact of digital technology on those using it [144].

Limitations

Considering the included studies, it is necessary to interpret the significant meta-analytic effect on the reduction of emotional experience in digital games with caution because of potential placebo effects, as discussed previously. In addition, only postintervention results were presented in the majority (28/39, 72%) of studies. This limited the ability to assess whether immediate improvements persisted and for how long. Hence, it is recommended that follow-up assessments should be conducted in large-scale studies that assess efficacy. Moreover, systematic review findings and subsequent discussions should be interpreted with caution because of the high risk of bias exhibited within the outcome measures. Researchers should endeavor to use validated and reliable acceptability and feasibility measures, and where this is not possible (eg, when obtaining nuanced qualitative information in iterative development workshops, web-based activities, focus groups, or interviews), a clear acknowledgment and explanation of the implications of using measures that may bias the outcomes should be provided. Finally, evident in this review is the limited number of large-scale RCTs. To push the emotion regulation digital intervention field forward, a transformation of the ethics and review board application process is required [141]. Currently, funding review panels frequently require highly detailed study protocols, with little to no consideration for the flexibility that is necessitated in collaborative design [141]. Encouraging greater flexibility in emotion regulation digital intervention development and evaluation plans may permit a

XSL•FO
RenderX

stronger research focus on vital acceptability and feasibility features and lead to the successful growth of this emerging field.

Although the findings discussed here reveal potential benefits of, and provide recommendations for, the rigorous progression of emotion regulation digital interventions in children and early adolescents, this systematic review and meta-analysis does include some limitations. The search strategy was broad, which may be seen as a strength at this early stage of the field's progression because it is imperative to understand the breadth of factors that may be implicated in its advancement. By contrast, this increased the number of required focal points of the review, which may have reduced its specificity. Furthermore, the focus on childhood and early adolescence is a strength—it enabled a nuanced understanding of this important developmental period. However, the meta-analysis did not include informant-reported effects. Although this decision was made to ensure homogeneity of the selected effect sizes, it might have limited the understanding of the benefits of emotion regulation digital intervention.

In conclusion, this review provides an important first step in the progression of emotion regulation digital interventions by synthesizing efficacy, feasibility, and acceptability data, published from 2008 to 2020, with a focus on childhood to early adolescence. The most consistent evidence came from digital games in the reduction of negative emotions, principally in children at risk of anxiety. However, variable methodologies, lack of follow-up assessment, and high risk of bias, inclusive of potential placebo effects within statistically influential neurofeedback-based digital game studies, limit definitive conclusions that may be made regarding the efficacy of such interventions. Engaging iterative intervention codevelopment with the sample who will eventually use the digital intervention and properly adjusting the difficulty to the intervention target is vital in achieving optimal acceptability and, specifically, addressing concerns around engagement. Finally, large-scale studies that assess emotion regulation as a key outcome using valid and reliable measures are urgently required to assess the extent to which emotion regulation ability may be improved in different samples of children and early adolescents through digital technology.

Acknowledgments

The Die Offene Tür (German for The Open Door) team of coinvestigators—Adam Barnard (playwright and theater director), JD (computer scientist), MM (medical doctor and researcher), BS (practicing and research psychiatrist), and KW (research psychologist)—grew during a sandpit event organized by the Open Innovation Center of Ludwig Boltzmann Society and facilitated by Know Innovation. The team would like to thank these organizations for their role in its formation. Die Offene Tür teamwork is jointly conceived and led by the coinvestigators. The team sees this as a fully meaningful interdisciplinary collaboration. This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (UIDB/04326/2020, UIDB/50021/2020, UIDP/04326/2020, and LA/P/0101/2020).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strings.
[DOCX File , 14 KB - games_v10i3e31456_app1.docx]

Multimedia Appendix 2
Master extraction table.
[XLSX File (Microsoft Excel File), 131 KB - games_v10i3e31456_app2.xlsx]

Multimedia Appendix 3
Risk-of-bias tool domains and definitions.
[DOCX File , 26 KB - games_v10i3e31456_app3.docx]

Multimedia Appendix 4
Intervention characteristic matrix summary.
[DOCX File , 30 KB - games_v10i3e31456_app4.docx]

Multimedia Appendix 5
Intervention characteristic summary.
[DOCX File , 19 KB - games_v10i3e31456_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

Risk-of-bias result summary.
[DOCX File , 66 KB - games_v10i3e31456_app6.docx]

Multimedia Appendix 7
Study and measures matrix.
[DOCX File , 29 KB - games_v10i3e31456_app7.docx]

Multimedia Appendix 8
Meta-analysis influence analysis summary.
[DOCX File , 14 KB - games_v10i3e31456_app8.docx]

Multimedia Appendix 9
Meta-analysis trim-and-fill funnel plot.
[DOCX File , 42 KB - games_v10i3e31456_app9.docx]

Multimedia Appendix 10
Intervention efficacy matrix.
[DOCX File , 28 KB - games_v10i3e31456_app10.docx]

Multimedia Appendix 11
Intervention feasibility matrix.
[DOCX File , 26 KB - games_v10i3e31456_app11.docx]

Multimedia Appendix 12
Intervention acceptability matrix.
[DOCX File , 20 KB - games_v10i3e31456_app12.docx]

References

1. Chervonsky E, Hunt C. Emotion regulation, mental health, and social wellbeing in a young adolescent sample: a concurrent and longitudinal investigation. Emotion 2019 Mar;19(2):270-282. [doi: 10.1037/emo0000432] [Medline: 29697988]
2. McLaughlin KA, Hatzenbuehler ML, Mennin DS, Nolen-Hoeksema S. Emotion dysregulation and adolescent psychopathology: a prospective study. Behav Res Ther 2011 Sep;49(9):544-554 [FREE Full text] [doi: 10.1016/j.brat.2011.06.003] [Medline: 21718967]
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
4. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022 Jan;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
5. Schriber RA, Guyer AE. Adolescent neurobiological susceptibility to social context. Dev Cogn Neurosci 2016 Jun;19:1-18 [FREE Full text] [doi: 10.1016/j.dcn.2015.12.009] [Medline: 26773514]
6. Steinberg L. Cognitive and affective development in adolescence. Trends Cogn Sci 2005 Feb;9(2):69-74. [doi: 10.1016/j.tics.2004.12.005] [Medline: 15668099]
7. Mitic M, Woodcock KA, Amering M, Krammer I, Stiehl KA, Zehetmayer S, et al. Toward an integrated model of supportive peer relationships in early adolescence: a systematic review and exploratory meta-analysis. Front Psychol 2021 Feb 25;12:589403 [FREE Full text] [doi: 10.3389/fpsyg.2021.589403] [Medline: 33716860]
8. Lamblin M, Murawski C, Whittle S, Fornito A. Social connectedness, mental health and the adolescent brain. Neurosci Biobehav Rev 2017 Sep;80:57-68. [doi: 10.1016/j.neubiorev.2017.05.010] [Medline: 28506925]
9. Birkeland MS, Breivik K, Wold B. Peer acceptance protects global self-esteem from negative effects of low closeness to parents during adolescence and early adulthood. J Youth Adolesc 2014 Jan;43(1):70-80 [FREE Full text] [doi: 10.1007/s10964-013-9929-1] [Medline: 23435859]
10. Thompson RA. Emotion and emotion regulation: two sides of the developing coin. Emotion Rev 2011 Jan 10;3(1):53-61. [doi: 10.1177/1754073910380969]
11. Sukhodolsky DG, Smith SD, McCauley SA, Ibrahim K, Piasecka JB. Behavioral interventions for anger, irritability, and aggression in children and adolescents. J Child Adolesc Psychopharmacol 2016 Feb;26(1):58-64 [FREE Full text] [doi: 10.1089/cap.2015.0120] [Medline: 26745682]

XSL•FO
RenderX

12. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
13. McRae K, Gross JJ. Emotion regulation. Emotion 2020 Feb;20(1):1-9. [doi: 10.1037/emo0000703] [Medline: 31961170]
14. Sheppes G, Suri G, Gross JJ. Emotion regulation and psychopathology. Annu Rev Clin Psychol 2015;11:379-405. [doi: 10.1146/annurev-clinpsy-032814-112739] [Medline: 25581242]
15. Gross JJ. The extended process model of emotion regulation: elaborations, applications, and future directions. Psychol Inq 2015 Mar 09;26(1):130-137. [doi: 10.1080/1047840X.2015.989751]
16. Schweizer S, Gotlib IH, Blakemore SJ. The role of affective control in emotion regulation during adolescence. Emotion 2020 Feb;20(1):80-86 [FREE Full text] [doi: 10.1037/emo0000695] [Medline: 31961183]
17. Spear LP. Rewards, aversions and affect in adolescence: emerging convergences across laboratory animal and human data. Dev Cogn Neurosci 2011 Oct;1(4):392-400 [FREE Full text] [doi: 10.1016/j.dcn.2011.08.001] [Medline: 21918675]
18. Riediger M, Wrzus C, Schmiedek F, Wagner GG, Lindenberger U. Is seeking bad mood cognitively demanding? Contra-hedonic orientation and working-memory capacity in everyday life. Emotion 2011 Jun;11(3):656-665. [doi: 10.1037/a0022756] [Medline: 21534659]
19. Ahmed SP, Bittencourt-Hewitt A, Sebastian CL. Neurocognitive bases of emotion regulation development in adolescence. Dev Cogn Neurosci 2015 Oct;15:11-25 [FREE Full text] [doi: 10.1016/j.dcn.2015.07.006] [Medline: 26340451]
20. Beck AT, Dozois DJ. Cognitive therapy: current status and future directions. Annu Rev Med 2011;62:397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
21. Abrams M, Ellis A. Rational emotive behaviour therapy in the treatment of stress. Br J Guid Couns 1994;22(1):39-50. [doi: 10.1080/03069889408253664]
22. Linehan MM. DBT Skills Training Manual. 2nd edition. New York, NY, USA: The Guilford Press; 2015.
23. James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev 2015 Feb 18;2015(2):CD004690 [FREE Full text] [doi: 10.1002/14651858.CD004690.pub4] [Medline: 25692403]
24. Chiles JA, Lambert MJ, Hatch AL. The impact of psychological interventions on medical cost offset: a meta-analytic review. Clin Psychol (New York) 1999;6(2):204-220. [doi: 10.1093/clipsy.6.2.204]
25. Moses T. Determinants of mental illness stigma for adolescents discharged from psychiatric hospitalization. Soc Sci Med 2014 May;109:26-34. [doi: 10.1016/j.socscimed.2013.12.032] [Medline: 24695363]
26. Wright A, Jorm AF, Mackinnon AJ. Labeling of mental disorders and stigma in young people. Soc Sci Med 2011 Aug;73(4):498-506. [doi: 10.1016/j.socscimed.2011.06.015] [Medline: 21794967]
27. Mukolo A, Heflinger CA, Wallston KA. The stigma of childhood mental disorders: a conceptual framework. J Am Acad Child Adolesc Psychiatry 2010 Feb;49(2):92-198 [FREE Full text] [doi: 10.1097/00004583-201002000-00003] [Medline: 20215931]
28. de Haan AM, Boon AE, de Jong JT, Hoeve M, Vermeiren RR. A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. Clin Psychol Rev 2013 Jul;33(5):698-711. [doi: 10.1016/j.cpr.2013.04.005] [Medline: 23742782]
29. Hoffmann JD, Brackett MA, Bailey CS, Willner CJ. Teaching emotion regulation in schools: translating research into practice with the RULER approach to social and emotional learning. Emotion 2020 Feb;20(1):105-109. [doi: 10.1037/emo0000649] [Medline: 31961187]
30. Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. Child Dev 2011;82(1):405-432. [doi: 10.1111/j.1467-8624.2010.01564.x] [Medline: 21291449]
31. Greenberg MT, Weissberg RP, O'Brien MU, Zins JE, Fredericks L, Resnik H, et al. Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. Am Psychol 2003;58(6-7):466-474. [doi: 10.1037/0003-066x.58.6-7.466] [Medline: 12971193]
32. Hajal NJ, Paley B. Parental emotion and emotion regulation: a critical target of study for research and intervention to promote child emotion socialization. Dev Psychol 2020 Mar;56(3):403-417. [doi: 10.1037/dev0000864] [Medline: 32077713]
33. England-Mason G, Gonzalez A. Intervening to shape children's emotion regulation: a review of emotion socialization parenting programs for young children. Emotion 2020 Feb;20(1):98-104. [doi: 10.1037/emo0000638] [Medline: 31961186]
34. Keppens G, Spruyt B. Truancy in Europe: does the type of educational system matter? Eur J Educ 2018 Aug 02;53(3):414-426. [doi: 10.1111/ejed.12282]
35. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
36. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]

XSL•FO
RenderX

37. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
38. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
39. Crone EA, Konijn EA. Media use and brain development during adolescence. Nat Commun 2018 Feb 21;9(1):588 [FREE Full text] [doi: 10.1038/s41467-018-03126-x] [Medline: 29467362]
40. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940 [FREE Full text] [doi: 10.1111/jcpp.13258] [Medline: 32572961]
41. Reynard S, Woodcock K. Digital interventions for emotion regulation in youth: a systematic literature review and meta-analysis. Prospero. 2018 Dec 11. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42018117608 [accessed 2018-12-11]
42. McDaid D, Park A, Wahlbeck K. The economic case for the prevention of mental illness. Annu Rev Public Health 2019 Apr 01;40:373-389. [doi: 10.1146/annurev-publhealth-040617-013629] [Medline: 30601725]
43. Veritas Health Innovation. Covidence Systematic Review Software. Covidence. URL: http://www.covidence.org/ [accessed 2020-07-01]
44. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
45. Wijnhoven LA, Creemers DH, Vermulst AA, Lindauer RJ, Otten R, Engels RC, et al. Effects of the video game 'Mindlight' on anxiety of children with an autism spectrum disorder: a randomized controlled trial. J Behav Ther Exp Psychiatry 2020 Sep;68:101548. [doi: 10.1016/j.jbtep.2020.101548] [Medline: 32155470]
46. Schoneveld EA, Malmberg M, Lichtwarck-Aschoff A, Verheijen GP, Engels RC, Granic I. A neurofeedback video game (MindLight) to prevent anxiety in children: a randomized controlled trial. Comput Human Behav 2016 Oct;63:321-333. [doi: 10.1016/j.chb.2016.05.005]
47. Schoneveld EA, Lichtwarck-Aschoff A, Granic I. Preventing childhood anxiety disorders: is an applied game as effective as a cognitive behavioral therapy-based program? Prev Sci 2018 Feb;19(2):220-232 [FREE Full text] [doi: 10.1007/s11121-017-0843-8] [Medline: 28956222]
48. Amon KL, Campbell A. Can children with AD/HD learn relaxation and breathing techniques through biofeedback video games. Aust J Educ Dev Psychol 2008;8:72-84.
49. Antle AN, Chesick L, Sridharan SK, Cramer E. East meets west: a mobile brain-computer system that helps children living in poverty learn to self-regulate. Pers Ubiquit Comput 2018 Jun 12;22(4):839-866. [doi: 10.1007/s00779-018-1166-x]
50. Beaumont R, Sofronoff K. A multi-component social skills intervention for children with Asperger syndrome: the Junior Detective Training Program. J Child Psychol Psychiatry 2008 Jul;49(7):743-753. [doi: 10.1111/j.1469-7610.2008.01920.x] [Medline: 18503531]
51. Beaumont R, Rotolone C, Sofronoff K. The secret agent society social skills program for children with high-functioning autism spectrum disorders: a comparison of two school variants. Psychol Schs 2015 Feb 19;52(4):390-402. [doi: 10.1002/pits.21831]
52. Beaumont RB, Pearson R, Sofronoff K. A novel intervention for child peer relationship difficulties: the secret agent society. J Child Fam Stud 2019 Jun 26;28(11):3075-3090. [doi: 10.1007/s10826-019-01485-7]
53. Carlier S, Van der Paelt S, Ongenae F, De Backere F, De Turck F. Empowering children with ASD and their parents: design of a serious game for anxiety and stress reduction. Sensors (Basel) 2020 Feb 11;20(4):966 [FREE Full text] [doi: 10.3390/s20040966] [Medline: 32054025]
54. Carroll A, Sanders O'Connor E, Houghton SJ, Hattie JA, Donovan C, Lynn S. A preliminary empirical evaluation of KooLKIDS: a school-based program to reduce early onset antisocial behaviour in children. J Intellect Dev Disabil 2016 Jun 12;42(1):22-35. [doi: 10.3109/13668250.2016.1190004]
55. Carroll A, McCarthy M, Houghton S, Sanders O'Connor E. Evaluating the effectiveness of KooLKIDS: an interactive social emotional learning program for Australian primary school children. Psychol Schs 2020 Feb 15;57(6):851-867. [doi: 10.1002/pits.22352]
56. Cohen Kadosh K, Luo Q, de Burca C, Sokunbi MO, Feng J, Linden DE, et al. Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. Neuroimage 2016 Jan 15;125:616-626 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.09.070] [Medline: 26475487]
57. David OA, Predatu RM, Cardo RA. A pilot study of the REThink online video game applied for coaching emotional understanding in children and adolescents in the therapeutic video game environment: the feeling better resources game. J Evid Based Psychother 2018 Mar 01;18(1):57-67. [doi: 10.24193/jebp.2018.1.5]

XSL•FO
RenderX

58. David OA, Cardo RA, Matu S. Is REThink therapeutic game effective in preventing emotional disorders in children and adolescents? Outcomes of a randomized clinical trial. Eur Child Adolesc Psychiatry 2019 Jan;28(1):111-122. [doi: 10.1007/s00787-018-1192-2] [Medline: 29992353]
59. David OA, Predatu R, Maffei A. REThink online video game for children and adolescents: effects on state anxiety and frontal alpha asymmetry. J Cogn Ther 2020 Jul 08;14(2):399-416. [doi: 10.1007/s41811-020-00077-4]
60. Einfeld SL, Beaumont R, Clark T, Clarke KS, Costley D, Gray KM, et al. School-based social skills training for young people with autism spectrum disorders. J Intellect Dev Disabil 2017 Jul 10;43(1):29-39. [doi: 10.3109/13668250.2017.1326587]
61. Filella G, Cabello E, Pérez-Escola N, Ros-Morente A. Evaluation of the emotional education program "happy 8-12" for the assertive resolution of conflicts among peers. Electron J Res Educ Psychol 2016 Dec 01;14(3):582-601. [doi: 10.14204/ejrep.40.15164]
62. Filella G, Ros-Morente A, Oriol X, March-Llanes J. The assertive resolution of conflicts in school with a gamified emotion education program. Front Psychol 2018 Nov 30;9:2353 [FREE Full text] [doi: 10.3389/fpsyg.2018.02353] [Medline: 30559691]
63. Goodman MS, Castro N, Sloan M, Sharma R, Widdowson M, Herrera E, et al. A neurovisceral approach to autism: targeting self-regulation and core symptoms using neurofeedback and biofeedback. Neuroregulation 2018 Mar 31;5(1):9-29. [doi: 10.15540/nr.5.1.9]
64. Heinrich H, Gevensleben H, Becker A, Rothenberger A. Effects of neurofeedback on the dysregulation profile in children with ADHD: SCP NF meets SDQ-DP - a retrospective analysis. Psychol Med 2020 Jan;50(2):258-263. [doi: 10.1017/S0033291718004130] [Medline: 30674360]
65. Houghton S, Carroll A, Zadow C, O'connor ES, Hattie J, Lynn S. Treating children with early-onset conduct problems and callous–unemotional traits: an empirical evaluation of KooLKIDS. Emot Behav Difficult 2017 Mar 15;22(4):350-364. [doi: 10.1080/13632752.2017.1301646]
66. Ip HH, Wong SW, Chan DF, Byrne J, Li C, Yuan VS, et al. Enhance emotional and social adaptation skills for children with autism spectrum disorder: a virtual reality enabled approach. Comput Educ 2018 Feb;117:1-15. [doi: 10.1016/j.compedu.2017.09.010]
67. Kahn J, Ducharme P, Rotenberg A, Gonzalez-Heydrich J. "RAGE-Control": a game to build emotional strength. Games Health J 2013 Feb;2(1):53-57. [doi: 10.1089/g4h.2013.0007] [Medline: 26196556]
68. Lackner N, Unterrainer HF, Skliris D, Shaheen S, Dunitz-Scheer M, Wood G, et al. EEG neurofeedback effects in the treatment of adolescent anorexia nervosa. Eat Disord 2016;24(4):354-374. [doi: 10.1080/10640266.2016.1160705] [Medline: 27027700]
69. Lutz B. An institutional case study: emotion regulation with HeartMath at Santa Cruz county children's mental health. Glob Adv Health Med 2014 Mar;3(2):68-71 [FREE Full text] [doi: 10.7453/gahmj.2014.007] [Medline: 24808984]
70. Rodriguez A, Rey B, Vara MD, Wrzesien M, Alcañiz M, Baños RM, et al. A VR-based serious game for studying emotional regulation in adolescents. IEEE Comput Graph Appl 2015;35(1):65-73. [doi: 10.1109/MCG.2015.8] [Medline: 25666600]
71. Rogel A, Loomis AM, Hamlin E, Hodgdon H, Spinazzola J, van der Kolk B. The impact of neurofeedback training on children with developmental trauma: a randomized controlled study. Psychol Trauma 2020 Nov;12(8):918-929. [doi: 10.1037/tra0000648] [Medline: 32658503]
72. Ruiz-Ariza A, Casuso RA, Suarez-Manzano S, Martínez-López EJ. Effect of augmented reality game Pokémon GO on cognitive performance and emotional intelligence in adolescent young. Comput Educ 2018 Jan;116:49-63. [doi: 10.1016/j.compedu.2017.09.002]
73. Scholten H, Malmberg M, Lobel A, Engels RC, Granic I. A randomized controlled trial to test the effectiveness of an immersive 3D video game for anxiety prevention among adolescents. PLoS One 2016 Jan 27;11(1):e0147763 [FREE Full text] [doi: 10.1371/journal.pone.0147763] [Medline: 26816292]
74. Schoneveld EA, Wols A, Lichtwarck-Aschoff A, Otten R, Granic I. Mental health outcomes of an applied game for children with elevated anxiety symptoms: a randomized controlled non-inferiority trial. J Child Fam Stud 2020 Jun 24;29(8):2169-2185. [doi: 10.1007/s10826-020-01728-y]
75. Schuurmans AA, Nijhof KS, Engels RC, Granic I. Using a videogame intervention to reduce anxiety and externalizing problems among youths in residential care: an initial randomized controlled trial. J Psychopathol Behav Assess 2018;40(2):344-354 [FREE Full text] [doi: 10.1007/s10862-017-9638-2] [Medline: 29937625]
76. Shum AK, Lai ES, Leung WG, Cheng MN, Wong HK, So SW, et al. A digital game and school-based intervention for students in Hong Kong: quasi-experimental design. J Med Internet Res 2019 Apr 05;21(4):e12003 [FREE Full text] [doi: 10.2196/12003] [Medline: 30950795]
77. Smith EN, Romero C, Donovan B, Herter R, Paunesku D, Cohen GL, et al. Emotion theories and adolescent well-being: results of an online intervention. Emotion 2018 Sep;18(6):781-788. [doi: 10.1037/emo0000379] [Medline: 29265838]
78. Sofronoff K, Silva J, Beaumont R. The secret agent society social-emotional skills program for children with a high-functioning autism spectrum disorder: a parent-directed trial. Focus Autism Other Dev Disabl 2017 Apr 28;32(1):55-70. [doi: 10.1177/1088357615583467]

XSL•FO
RenderX

79. Torrado JC, Gomez J, Montoro G. Emotional self-regulation of individuals with autism spectrum disorders: smartwatches for monitoring and interaction. Sensors (Basel) 2017 Jun 11;17(6):1359 [FREE Full text] [doi: 10.3390/s17061359] [Medline: 28604607]
80. Vara MD, Baños RM, Rasal P, Rodríguez A, Rey B, Wrzesien M, et al. A game for emotional regulation in adolescents: the (body) interface device matters. Comput Human Behav 2016 Apr;57:267-273. [doi: 10.1016/j.chb.2015.12.033]
81. Vara MD, Baños RM, Rasal P, Rodríguez A, Rey B, Wrzesien M, et al. A VR-based serious game to regulate joy in adolescents: a comparison of different devices. In: Proceedings of the International Summit on eHealth 360°. 2016 Presented at: eHealth '16; June 14-16, 2016; Budapest, Hungary p. 135-142. [doi: 10.1007/978-3-319-49655-9_18]
82. Wrzesien M, Rodríguez A, Rey B, Alcañiz M, Baños RM, Vara MD. How the physical similarity of avatars can influence the learning of emotion regulation strategies in teenagers. Comput Human Behav 2015 Feb;43:101-111. [doi: 10.1016/j.chb.2014.09.024]
83. Yuan SN, Ip HH. Using virtual reality to train emotional and social skills in children with autism spectrum disorder. London J Prim Care (Abingdon) 2018 Jun 7;10(4):110-112 [FREE Full text] [doi: 10.1080/17571472.2018.1483000] [Medline: 30083244]
84. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
85. Risk of Bias online library. Cochrane Methods Webpages. 2018. URL: https://methods.cochrane.org/ [accessed 2018-08-31]
86. Williams M. Feasibility and pilot studies: a guide for NIHR Research Design Service advisors. NHS National Institute for Health Research. 2016. URL: https://www.rds-eastmidlands.nihr.ac.uk/ [accessed 2018-08-31]
87. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, International Stroke Trial Collaborative Group, European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27):iii-173 [FREE Full text] [doi: 10.3310/hta7270] [Medline: 14499048]
88. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Lancaster University. 2006. URL: https://tinyurl.com/56xtp86h [accessed 2018-08-31]
89. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2017. URL: https://www.R-project.org/ [accessed 2020-11-23]
90. Lüdecke D. esc: Effect Size Computation for Meta Analysis (Version 0.5.1). The Comprehensive R Archive Network. 2019. URL: https://cran.r-project.org/web/packages/esc/index.html [accessed 2020-11-23]
91. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.1. London, UK: Cochrane Collaboration; 2020.
92. Wang Z, Nayfeh T, Sofiyeva N, Ponte OJ, Rajjoub R, Malandris K, et al. Including non-inferiority trials in contemporary meta-analyses of chronic medical conditions: a meta-epidemiological study. J Gen Intern Med 2020 Jul;35(7):2162-2166 [FREE Full text] [doi: 10.1007/s11606-020-05821-2] [Medline: 32318902]
93. Davey J, Turner RM, Clarke MJ, Higgins JP. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol 2011 Nov 24;11:160 [FREE Full text] [doi: 10.1186/1471-2288-11-160] [Medline: 22114982]
94. López-López JA, Van den Noortgate W, Tanner-Smith EE, Wilson SJ, Lipsey MW. Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: a Monte Carlo simulation. Res Synth Methods 2017 Dec;8(4):435-450. [doi: 10.1002/jrsm.1245] [Medline: 28556477]
95. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
96. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Nov;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
97. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
98. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis in R: A Hands-on Guide. Boca Raton, FL, USA: Chapmann & Hall/CRC Press; 2021.
99. Cuijpers P. Meta-Analyses in Mental Health Research: A Practical Guide. Amsterdam, The Netherlands: Vrije Universiteit Amsterdam; 2017.
100. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016 Mar;7(1):55-79 [FREE Full text] [doi: 10.1002/jrsm.1164] [Medline: 26332144]
101. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Mahwah, NJ, USA: Lawrence Earlbaum Associates; Aug 12, 1988.
102. Hedges LV. Distribution theory for glass's estimator of effect size and related estimators. J Educ Stat 1981;6(2):107-128. [doi: 10.3102/10769986006002107]
103. Hedges L, Olkin I. Statistical Models for Meta-Analysis. New York, NY, USA: Academic Press; Jul 10, 1985.

XSL•FO
RenderX

104. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; 2011.
105. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
106. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
107. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
108. Rothstein HR, Sutton AJ, Borenstein M. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Hoboken, NJ, USA: John Wiley & Sons; 2006.
109. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008 Oct;61(10):991-996. [doi: 10.1016/j.jclinepi.2007.11.010] [Medline: 18538991]
110. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
111. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
112. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
113. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012 Oct 15;22(3):276-282. [doi: 10.11613/BM.2012.031]
114. Sweller J, Ayres P, Kalyuga S. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies. Cognitive Load Theory. New York, NY, USA: Springer; 2011.
115. Wouters P, Paas F, van Merriënboer JJ. How to optimize learning from animated models: a review of guidelines based on cognitive load. Rev Educ Res 2008 Sep 1;78(3):645-675. [doi: 10.3102/0034654308320320]
116. Catalano CE, Luccini AM, Mortara M. Best practices for an effective design and evaluation of serious games. Int J Serious Games 2014 Feb 03;1(1):1-13. [doi: 10.17083/ijsg.v1i1.8]
117. Whyte EM, Smyth JM, Scherf KS. Designing serious game interventions for individuals with autism. J Autism Dev Disord 2015 Dec;45(12):3820-3831. [doi: 10.1007/s10803-014-2333-1] [Medline: 25488121]
118. Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev Gen Psychol 2010 Jun 01;14(2):154-166. [doi: 10.1037/a0019440]
119. Ávila-Pesántez D, Rivera LA, Alban MS. Approaches for serious game design: a systematic literature review. Comput Educ 2017 Sep;8(3):1-11.
120. Guendelman S, Medeiros S, Rampes H. Mindfulness and Emotion Regulation: Insights from Neurobiological, Psychological, and Clinical Studies. Front Psychol 2017 Mar 06;8:220 [FREE Full text] [doi: 10.3389/fpsyg.2017.00220] [Medline: 28321194]
121. Fovet T, Micoulaud-Franchi JA, Vialatte FB, Lotte F, Daudet C, Batail JM, et al. On assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms? Brain 2017 Oct 01;140(10):e63. [doi: 10.1093/brain/awx211] [Medline: 28969378]
122. Schabus M. Reply: on assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms? Brain 2017 Oct 01;140(10):e64 [FREE Full text] [doi: 10.1093/brain/awx212] [Medline: 28969379]
123. Schabus M, Griessenberger H, Gnjezda MT, Heib DP, Wislowska M, Hoedlmoser K. Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. Brain 2017 Apr 01;140(4):1041-1052 [FREE Full text] [doi: 10.1093/brain/awx011] [Medline: 28335000]
124. Tsai N, Buschkuehl M, Kamarsu S, Shah P, Jonides J, Jaeggi SM. (Un)great expectations: the role of placebo effects in cognitive training. J Appl Res Mem Cogn 2018 Dec;7(4):564-573 [FREE Full text] [doi: 10.1016/j.jarmac.2018.06.001] [Medline: 31660288]
125. Zhou Y, Daukantaitė D, Lundh LG, Wångby-Lundh M, Ryde A. Adolescents' emotion regulation strategies questionnaire: initial validation and prospective associations with nonsuicidal self-injury and other mental health problems in adolescence and young adulthood in a Swedish youth cohort. Front Psychiatry 2020 May 28;11:462 [FREE Full text] [doi: 10.3389/fpsyt.2020.00462] [Medline: 32547430]
126. Ferschmann L, Vijayakumar N, Grydeland H, Overbye K, Mills KL, Fjell AM, et al. Cognitive reappraisal and expressive suppression relate differentially to longitudinal structural brain development across adolescence. Cortex 2021 Mar;136:109-123 [FREE Full text] [doi: 10.1016/j.cortex.2020.11.022] [Medline: 33545616]
127. McRae K, Gross JJ, Weber J, Robertson ER, Sokol-Hessner P, Ray RD, et al. The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. Soc Cogn Affect Neurosci 2012 Jan;7(1):11-22 [FREE Full text] [doi: 10.1093/scan/nsr093] [Medline: 22228751]

XSL•FO
RenderX

128. Messina I, Bianco S, Sambin M, Viviani R. Executive and semantic processes in reappraisal of negative stimuli: insights from a meta-analysis of neuroimaging studies. Front Psychol 2015 Jul 13;6:956 [FREE Full text] [doi: 10.3389/fpsyg.2015.00956] [Medline: 26217277]
129. Burrell TL, Borrego Jr J. Parents' involvement in ASD treatment: what is their role? Cogn Behav Pract 2012 Aug;19(3):423-432. [doi: 10.1016/j.cbpra.2011.04.003]
130. Girolametto L, Tannock R. Correlates of directiveness in the interactions of fathers and mothers of children with developmental delays. J Speech Hear Res 1994 Oct;37(5):1178-1191. [doi: 10.1044/jshr.3705.1178] [Medline: 7823559]
131. Jouen AL, Narzisi A, Xavier J, Tilmont E, Bodeau N, Bono V, MICHELANGELO Study Group. GOLIAH (Gaming Open Library for Intervention in Autism at Home): a 6-month single blind matched controlled exploratory study. Child Adolesc Psychiatry Ment Health 2017 Mar 22;11:17 [FREE Full text] [doi: 10.1186/s13034-017-0154-7] [Medline: 28344643]
132. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
133. Fleming T, Merry S, Stasiak K, Hopkins S, Patolo T, Ruru S, et al. The importance of user segmentation for designing digital therapy for adolescent mental health: findings from scoping processes. JMIR Ment Health 2019 May 08;6(5):e12656 [FREE Full text] [doi: 10.2196/12656] [Medline: 31066705]
134. Hermann A, Kress L, Stark R. Neural correlates of immediate and prolonged effects of cognitive reappraisal and distraction on emotional experience. Brain Imaging Behav 2017 Oct;11(5):1227-1237. [doi: 10.1007/s11682-016-9603-9] [Medline: 27709512]
135. Cludius B, Mennin D, Ehring T. Emotion regulation as a transdiagnostic process. Emotion 2020 Feb;20(1):37-42. [doi: 10.1037/emo0000646] [Medline: 31961175]
136. Davis EL, Levine LJ. Emotion regulation strategies that promote learning: reappraisal enhances children's memory for educational information. Child Dev 2013;84(1):361-374. [doi: 10.1111/j.1467-8624.2012.01836.x] [Medline: 22906113]
137. English T, John OP, Srivastava S, Gross JJ. Emotion regulation and peer-rated social functioning: a four-year longitudinal study. J Res Pers 2012 Dec 01;46(6):780-784 [FREE Full text] [doi: 10.1016/j.jrp.2012.09.006] [Medline: 23471162]
138. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]
139. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
140. Robb N, Waller A, Woodcock KA. Developing a task switching training game for children with a rare genetic syndrome linked to intellectual disability. Simul Gaming 2019 Mar 06;50(2):160-179. [doi: 10.1177/1046878119834319]
141. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018 Oct 18;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
142. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
143. Hollis C, Livingstone S, Sonuga-Barke E. Editorial: the role of digital technology in children and young people's mental health - a triple-edged sword? J Child Psychol Psychiatry 2020 Aug;61(8):837-841. [doi: 10.1111/jcpp.13302] [Medline: 32706126]
144. Orben A. The Sisyphean cycle of technology panics. Perspect Psychol Sci 2020 Sep;15(5):1143-1157 [FREE Full text] [doi: 10.1177/1745691620919372] [Medline: 32603635]

Abbreviations

ASD: autism spectrum disorder
CBT: cognitive behavior therapy
EEG: electroencephalogram
fMRI: functional magnetic resonance imaging
HRV: heart rate variability
NHS: National Health Service
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by N Zary; submitted 23.06.21; peer-reviewed by A Daros, M Rus-Calafell, Z Aghaei; comments to author 16.03.22; revised version received 02.05.22; accepted 27.05.22; published 19.08.22.

Please cite as:
Reynard S, Dias J, Mitic M, Schrank B, Woodcock KA
Digital Interventions for Emotion Regulation in Children and Early Adolescents: Systematic Review and Meta-analysis
JMIR Serious Games 2022;10(3):e31456
URL: https://games.jmir.org/2022/3/e31456
doi:10.2196/31456
PMID:35984681

©Sally Reynard, Joao Dias, Marija Mitic, Beate Schrank, Kate Anne Woodcock. Originally published in JMIR Serious Games (https://games.jmir.org), 19.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Serious Games Supporting the Prevention and Treatment of Alcohol and Drug Consumption in Youth: Scoping Review

Juan Martínez-Miranda[1*], PhD; Ismael Edrein Espinosa-Curiel[1*], PhD

Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad de Transferencia Tecnológica Tepic, Tepic, Mexico
[*]all authors contributed equally

Corresponding Author:
Juan Martínez-Miranda, PhD
Centro de Investigación Científica y de Educación Superior de Ensenada
Unidad de Transferencia Tecnológica Tepic
Andador 10, Ciudad del Conocimiento
Tepic, 63173
Mexico
Phone: 52 3111295930 ext 28602
Email: jmiranda@cicese.mx

Abstract

Background: The consumption of alcohol and drugs, particularly in adolescents and young adults, has increased worldwide in the last several years, representing a significant public health challenge. Serious games have the potential to support preventive and treatment interventions for substance use, facilitating the acquisition of relevant knowledge and the motivation for changes in attitudes and behaviors regarding substance consumption.

Objective: This scoping review aims to analyze a set of 7 relevant characteristics of current serious games designed to support the prevention and treatment of alcohol and drug consumption in adolescents and young adults—the substance addressed, the type of intervention, the theoretical basis, the computational techniques used, the mechanism for data security and privacy, the evaluation procedure followed, and the main results obtained.

Methods: The review was performed by following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Data were retrieved from January 2010 to May 2022, using PubMed, Scopus (Elsevier), IEEE Xplore, and ACM Digital as data sources. The eligibility criteria included studies that described serious games designed to support the prevention or treatment of alcohol and drug consumption, targeted a population aged between 12 and 30 years, and included an evaluation procedure. Authors (JMM and IEEC) individually screened the titles and abstracts, and then full articles were reviewed for a final inclusion decision.

Results: A total of 629 records were obtained, and 29 (4.6%) fulfilled the inclusion criteria. Most of the serious games (14/29, 48%) were focused on the prevention or treatment of alcohol use. The type of intervention that was the most supported was prevention (18/29, 62%), and most studies mentioned the theory, theoretical construct, or therapeutic technique used as a foundation (22/29, 76%). Most of the studies only provided information about the platform for execution (23/29, 79%), and few described the use of computational techniques, such as virtual reality or motion-based interaction (5/29, 17%). A small set of studies (10/29, 34%) explicitly mentioned how data security and privacy were addressed. Most of the reported evaluation protocols were pilot studies (11/29, 38%), followed by randomized controlled trials (10/29, 34%), and the reported results were positive in terms of acceptability, usability, and efficacy. However, more research is needed to assess long-term effects.

Conclusions: Given the increasing interest in the use of serious games as digital interventions to support the prevention or treatment of substance use, knowing their main features is highly important. This review highlights whether and how current serious games incorporate 7 key features that are useful to consider for the further development of the area.

(JMIR Serious Games 2022;10(3):e39086) doi:10.2196/39086

KEYWORDS

serious games; substance use; alcohol and drugs; young population; mobile phone

XSL•FO
RenderX

Introduction

Background

The use of illicit drugs and alcohol is a serious public health problem worldwide that affects an important percentage of the population. The United Nations World Drug Report 2021 estimates that approximately 5.5% of the population aged between 15 and 64 years have used drugs at least once in the past year, whereas 36.3 million people (13% of the total number of people who use drugs) have drug use disorders [1]. Furthermore, the World Health Organization (WHO) estimates that the total alcohol consumption per capita in the world's population aged ≥15 years is at the level of 6.4 L of pure alcohol, and it is expected to increase to 7.0 L in 2025 [2].

Adolescence is the developmental period with the highest risk of developing a substance use disorder [3]. Studies on brain development indicate that this period of life is characterized by suboptimal decisions and actions that are associated with risk behaviors, including an increased incidence of substance abuse [4]. The health consequences of alcohol and drugs in this population are highly negative as the consumption of substances is associated with neurocognitive alterations that can lead to behavioral, emotional, social, and academic problems later in life [5]. Moreover, some studies show that people of younger ages are disproportionately affected by alcohol when compared with individuals of older ages. The proportion of all deaths attributable to alcohol consumption is most significant among those aged 20 to 39 years, representing 13.5% of all deaths among these people [2]. Furthermore, drug overdose is behind the deaths of many adolescents, which have increased in the last decades, mainly because of the consumption of opioids [6]. Thus, interventions for preventing and treating alcohol and drug consumption in youth are necessary to tackle substance abuse to minimize the negative consequences in older ages. These interventions are crucial during adolescence and emerging adulthood because of evidence that precocity has been associated with a further increased risk of developing substance use disorders [7].

Different intervention strategies are currently implemented to mitigate substance use in adolescents and young adults, including universal prevention, selective prevention, and treatment [8]. Although the evidence shows that these interventions are effective, there is also a critical need to increase the availability and accessibility of these health services to a greater population [9]. Owing to the wide use and acceptance of technology among the young population, different web-based applications and mobile apps have been developed to support and complement traditional interventions or are even provided as stand-alone interventions. These digital applications offer some advantages such as minimal cost [10], reducing the burden on health care professionals [11], and the personalization of the intervention to different individuals [12]. In addition to web-based applications and mobile apps, one of the digital solutions that would offer different advantages to support substance use interventions, particularly for adolescents and young adults, is the development of serious games.

Serious games are computer games developed with objectives not only related to users' entertainment but also to facilitate the increase of skills and abilities, gain knowledge, or acquire experience. The use of serious games for health has increased in the last several years because of their capabilities to simulate real-life situations, collect information that helps identify specific conditions and behaviors, and provide information to involve and support the user at different stages of the health care process [13]. Serious games have been used as digital-based interventions for different health care problems such as the prevention of obesity in children [14], supporting children with chronic diseases [15], promoting the treatment of mental illness [16], and providing information about how to prevent COVID-19 [17], to name a few.

Substance use prevention and treatment are also benefiting from serious games. Different research efforts are currently dedicated to developing and evaluating serious games to facilitate the gain of knowledge and the promotion of behavior change in individuals facing health problems related to substance use. Thus, it is essential to know how these solutions are currently designed, implemented, and evaluated in terms of the following questions: (1) What are the substances whose use is intended to be prevented or treated with the support of serious games? (2) How are these serious games designed and used (as stand-alone interventions or as part of a prevention or treatment program)? (3) What is the theoretical background on which these serious games are based? (4) What are the main computer-based techniques and methods (eg, augmented reality, artificial intelligence, and brain-computer interfaces) implemented in these serious games? (5) What are the types of evaluation and what are the main outcomes assessed with these serious games? (6) What mechanism is implemented to assure the security and confidentiality of users' data? (7) What are the reported efficacy and limitations of the serious games used for the prevention or treatment of substance use?

Objectives

To our knowledge, there are 2 previous reviews addressing the use of serious games for the prevention of alcohol and drug use, but both are exclusively focused on educational purposes [18,19]. Thus, the objective of this study was to perform a scoping review of serious games used for the prevention and treatment of alcohol and other drug use considering not only educational or learning objectives but also the promotion of behavior change. By answering the aforementioned 7 questions, this review will contribute to a better understanding of how serious games are currently designed, used, and evaluated, as well as of the reported impact they have as digital interventions for the prevention and treatment of substance use in adolescents and young adults.

Methods

We followed a scoping review methodology to synthesize concepts and research concerning the use of serious games as interventions designed to support the prevention and treatment of alcohol and drug consumption in youth. This protocol followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping

XSL•FO
RenderX

Reviews) methodology [20] to ensure that our review was conducted systematically and was bias-free.

Eligibility Criteria

The studies included in this scoping review were studies published between January 2010 and May 2022 that focused on the description of (1) *serious video games* developed with the objective to (2) *support the prevention or treatment* of (3) *alcohol and drug use*, targeted (4) *a population aged between 12 and 30 years*, and included (5) an *evaluation procedure.*

Serious Games

To include only serious video games and exclude other types of interactive game-like applications (eg, applications implementing simulated virtual environments or conversational synthetic characters), the main inclusion criterion was to identify in the retrieved works the five components of a serious game described in the study by Wattanasoontorn et al [13]: (1) implements rule or gameplay, (2) contains a challenge, (3) implements some type of user interaction, (4) has an explicit objective (entertainment), and (5) has an implicit objective (increasing skills and abilities, gaining knowledge, or acquiring experience).

Drugs

When referring to drugs in this study, we adopt the WHO definition, as follows: psychoactive drugs that, when taken in or administered into one's system, affect mental processes (eg, perception, consciousness, cognition, or mood and emotions). Psychoactive drugs belong to a broader category of psychoactive substances that also include alcohol and nicotine [21]. As we are also interested in substances that, if not identified and treated adequately, would cause mental health problems, in this review we excluded studies focused only on the prevention or treatment of tobacco use.

Adolescents and Youth Population

Owing to the evidence that adolescents and young adults are particularly affected by alcohol and drug consumption, we focused this review on serious games designed for and evaluated in this population. Nevertheless, there are no clear guidelines for determining what ages should be included in the designation of young adulthood. The WHO categorizes young people as adolescents and young adults from 10 through 24 years of age, and the United Nations defines youth as 15 to 24 years of age [22]. We decided to include all the studies considering the range from 12 to 30 years in line with other research that consider this age range when referring to adolescents and young adults in studies related to alcohol and drug consumption [23-25].

Thus, articles were excluded if they described studies (1) that did not describe a serious game or were not related to substance use, (2) whose main objective was not the prevention or treatment of substance use, (3) that did not include adolescents and young adults as the target population (age: range 12-30 years), (4) that did not include an evaluation procedure, (5) that described literature reviews, (6) that were focused only on the prevention or treatment of tobacco use, (7) whose main aim was the identification or screening of individuals at risk of substance use, (8) that were not in English, (9) that were repeated, and (10) that were not research articles.

Information Sources

PubMed, Scopus (Elsevier), IEEE Xplore, and ACM Digital were used to search for published papers. These 4 databases cover medical and computer science literature, allowing for comprehensive topic and field searches. The first search concluded in October 2021, and the final search concluded in May 2022. Some papers cited in the retrieved articles (hand searched) were also considered to complete the search.

Search

Depending on each database, the specific syntax of the queries was different but, in all cases, the same words were used to represent the constructs of (1) serious games together with (2) substance use but excluding (3) game and internet addiction. The query used was integrated with the following words: ("serious games" OR "gamification" OR "gamified" OR "games for health" OR "educational game" OR "videogame*" OR "game-based" OR "video game*") AND ("substance use" OR "substance abuse" OR "substance addiction" OR "alcohol addiction" OR "alcohol use" OR "alcohol abuse" OR "drug use" OR "drug abuse" OR "drug addiction") AND NOT ("game addiction" OR "gaming disorder" OR "gambling" OR "gaming problems" OR "smartphone addiction" OR "internet addiction" OR "video gaming").

Study Selection

A multistage screening process was conducted. First, 2 reviewers (JMM and IEEC) performed title and abstract screening. Articles that both researchers included then entered a second phase, where full texts were reviewed for the final inclusion decision. When some papers were related to the same study or application, the most recent one was selected unless significant differences were reported in the evaluation protocol (eg, the inclusion of different outcomes or testing with a different target population). The review of the articles and the data extraction were carried out separately, and any disagreements were resolved through discussion until a consensus was reached.

Data Charting and Synthesis of Results

All the studies meeting the inclusion criteria were selected for the review, and the data extracted were those that allowed for the answering of the seven questions listed in the Introduction section: (1) the type of substance use (alcohol or other drugs) the serious game aimed to prevent or treat, (2) the type of intervention that the serious game supported (as a stand-alone application or as part of a prevention or treatment program), (3) the theoretical basis used as the background model of the serious game, (4) the main computer-based methods or techniques implemented in the serious game, (5) the type of evaluation and procedure, (6) the mechanism implemented to maintain the security or confidentiality of users' data, and (7) the reported results in terms of the different outcomes relevant to the prevention or treatment of substance use (eg, gain of knowledge, skill development, or behavior change).

The authors developed, calibrated, and used a template containing different sections to extract and summarize the

XSL•FO
RenderX

aforementioned data. We further describe the main findings that emerged across the studies.

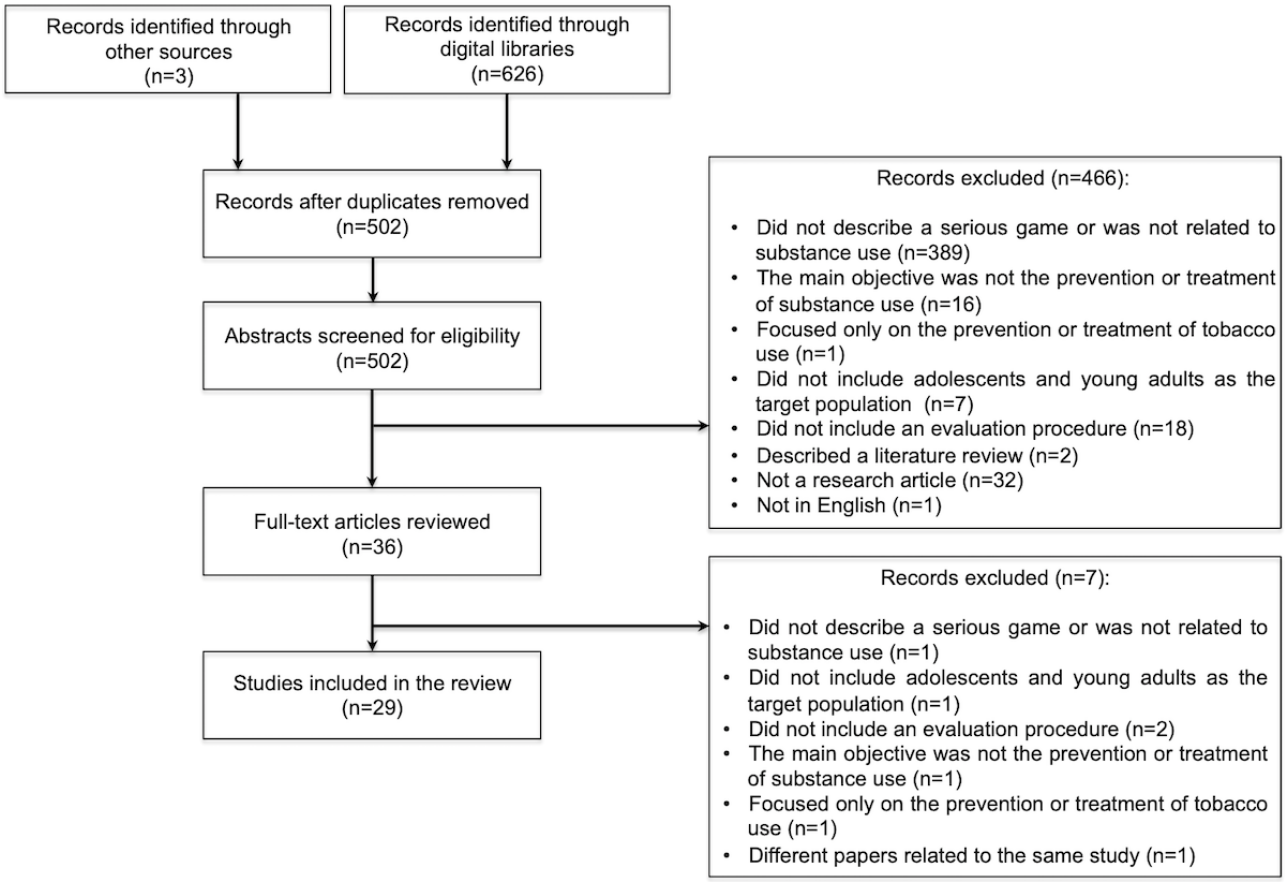
Results

Overview

A total of 629 records were identified from the searching process. From the 4 digital libraries, 99.5% (626/629) of the records were retrieved, and an additional 0.5% (3/629) of papers were obtained through hand searching. The records retrieved from each digital library were as follows: 21.1% (133/629) from PubMed, 33.1% (208/629) from Scopus, 26.4% (166/629) from IEEE Xplore, and 18.9% (119/629) from ACM Digital. In the first stage, after removing all the duplicated records, a total of 502 papers were screened for eligibility. After reading the titles and abstracts, 92.8% (466/502) of the records were discarded based on the exclusion criteria. A total of 36 articles were full-text reviewed, and 7 (19%) were excluded after the review. As a result, 29 studies were considered for further analysis. Figure 1 presents the flow diagram of the different review phases.

Most of the 29 studies were conducted in North America (United States and Canada; n=16, 55%), followed by studies developed in Europe (n=6, 21%), Australia (n=3, 10%), Brazil (n=2, 7%), and the Philippines (n=2, 7%). Most of the reviewed serious games (14/29, 48%) were developed to address problems associated with the consumption of alcohol. The type of intervention mainly supported was prevention (18/29, 62%), and the type of evaluation protocol most reported was pilot study (11/29, 38%) followed by randomized controlled trial (RCT; 10/29, 34%). In terms of the computational techniques used, most of the reviewed serious games only reported the type of platform where the game could be executed: web-based (11/29, 38%), mobile-based (7/29, 24%), multi-platform (3/29, 10%), or PC (2/29, 7%). Table 1 presents details of the general results. The following subsections describe the main findings to answer the 7 defined questions in detail.

Figure 1. Flow diagram of the review phases indicating the reasons for exclusion.



XSL•FO
RenderX

Table 1. Overview of the general results related to some of the assessed characteristics of the serious games (N=29).

Characteristic	Studies, n (%)
Type of substance addressed	
Alcohol	14 (48)
Drugs	8 (28)
Alcohol and drugs	7 (24)
Type of intervention	
Supporting prevention	18 (62)
Supporting treatment	11 (38)
Evaluation protocol	
Pilot study	11 (38)
RCT[a]	10 (34)
Quasi-experimental study	5 (17)
Comparative study	2 (7)
Proof-of-concept study	1 (3)
Computational techniques used	
Web-based	11 (38)
Mobile-based	7 (24)
Multi-platform	3 (10)
Use of virtual reality	3 (10)
Use of Kinect	2 (7)
PC-based	2 (7)
Not specified	1 (3)

[a]RCT: randomized controlled trial.

Type of Substance Addressed

In terms of the substances addressed, of the 29 studies, 13 (45%) were focused on problems associated with the consumption or abuse of alcohol, 7 (24%) were focused on drugs (such as methamphetamine, inhalants, cannabis, ecstasy, opioids, lysergic acid diethylamide, cocaine, and heroin), and 7 (24%) addressed problems for both alcohol and drug use. It is important to note that 7% (2/29) of the studies considered comorbidity—of these 2 studies, 1 (50%) focused on alcohol use disorder and depression [26,27], and 1 (50%) focused on substance use and relationship violence [28].

Type of Intervention

Most of the evaluated serious games (18/29, 62%) were developed to support the prevention of substance use, whereas the rest of the serious games (11/29, 38%) were focused on the enhancement of the provided treatment or on supporting individuals during specific stages of their treatment. All the serious games designed with prevention objectives (18/29, 62%) implemented their games' mechanics to facilitate the increase of knowledge on different topics, such as the negative consequences of nonmedical prescription drug use [29], the biological consequences of the abuse of inhalants [30], how the brain is affected by alcohol and other drugs [31,32], how to identify and reduce risky health behaviors [33,34] as well as carry out protective behavioral techniques [35,36], and the effects and consequences of alcohol and drug consumption [37,38].

Complementarily, the serious games focused on treatment (11/29, 38%) were designed to complement the therapeutic process for individuals with addiction problems or at risk and minimize relapse [39,40], reduce consumption [41-44], support recovery efforts [45], and promote a better adherence to the treatment [26,27]. Regarding the design of use, most of the serious games developed with prevention objectives (18/29, 62%) were stand-alone applications, and just a few (4/13, 31%) were designed as part of school-based prevention courses or curricula where teachers or social workers acted as facilitators [28,29,32,36,46]. On the contrary, most of the serious games developed with objectives to support treatment (11/29, 38%) were designed as technological-based tools to be used under the supervision of treatment providers.

Theoretical Background

The theoretical backgrounds used as the foundation of the analyzed serious games were varied in those studies that explicitly mentioned their theoretical basis (22/29, 76%). A total of 14 studies (14/29, 48%) referred either to a specific theory or theories, a theoretical construct, or an intervention technique. Only a few studies (8/29, 28%) referred to a theory jointly with the particular theoretical constructs or intervention

XSL•FO
RenderX

techniques used in the serious game [33,40-44,47,48]. Many of the features designed in the serious games for knowledge increasing and the fostering of health behavior adoption were based on the theory of reasoned action [29,47] or the social cognitive theory (SCT) alone [33,37] or in combination with other theories such as the reinforcement theory of motivation [48], the multiple intelligences theory [35], or the protection motivation theory [34]. The serious games with a focus on treatment used theories, theoretical constructs, or intervention techniques for behavior change, such as the I-Change Model [41], coping skill training [39], cognitive bias modification of attention [49], cognitive behavioral therapy [27,45], or cue exposure therapy [40]. A total of 24% (7/29) of the studies, all with prevention objectives, did not mention what theory, theoretical construct, or intervention technique was implemented in the serious game. Table 2 presents in detail the theoretical basis reported in each study split into theory and theoretical construct or intervention technique.

XSL•FO
RenderX

Table 2. Summary of the main characteristics of the reviewed serious games (N=29).

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Cheng et al [31]—Drugs and the Brain	Methamphetamine	Prevention—education on the impact of methamphetamine abuse on the brain	Not specified	Not specified	Virtual reality learning environment	Pilot study (pre- and posttest measurements) with 175 visitors (aged 6 to 82 years) to a museum	Not specified	Improvement of knowledge of basic neuroscience concepts and understanding and attitudes toward the impact of methamphetamine abuse.
Klisch et al [30]—Uncommon Scents	Inhalants (toxic chemicals)	Prevention—education on the biological consequences and risk of inhaling toxic chemicals	Not specified	Constructive learning	Web-based serious game	Pilot study (pre- and posttest measurements) involving 444 middle school students (sixth-, seventh-, and eighth-graders)	Data collected anonymously	Significant gains in science content knowledge were obtained. A shift to more negative attitudes toward inhalants was also observed. The most negative shift was among eighth-grade students. Posttest knowledge gains were the strongest predictor of attitude change across all grade levels.
Klisch et al [29]—Bitter Pill and Fatal Interactions	Prescription drugs	Prevention—education on the risk of prescription drug abuse	Theory of reasoned action	Not specified	Web-based serious game	Pilot study (pre- and posttest measurements) with 179 high school students (11th and 12th grade) divided into 2 groups assigning each group to a different game (Bitter Pill and Fatal Interactions)	Data collected anonymously	In both groups, the attitudes toward prescription drug abuse became significantly more negative. A moderate effect size in the Bitter Pill group and a large effect size in the Fatal Interactions group were found.
Sánchez and Bartel [39]—Arise	Alcohol and drugs	Treatment—education and building of coping skills for relapse prevention	Not specified	Coping skill training	Web-based serious game	Feasibility study (posttest measurements) with 8 treatment providers and a pilot study (posttest measurements) with 9 adolescents in substance abuse treatment	Data collected anonymously	The treatment providers rated the serious game with high likability and usefulness. Adolescents rated the game as easy to use, highly acceptable, and very useful as a substance abuse relapse prevention tool.
Elias-Lambert et al [28]—Choices & Consequences	Not explicitly mentioned; the study only refers to "substance abuse" and relationship violence	Prevention—education on the substance abuse and relationship violence challenges, the possible actions, and the consequences associated with those actions	Not specified	Situated experiential learning	Multiuser, mobile-based serious game	Exploratory proof of concept (posttest measurements; 6 focus groups) with 44 youth school students (aged >14 years)	Not specified	Participants reported enjoyment related to the visual and auditory aspects of the game and the facts provided to give further information about selected actions and their consequences. In terms of engagement, some of the participants reported that the scenarios were realistic and could connect and engage with them. Other feedback indicates game preference as a learning tool over traditional formats.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Jander et al [41]—What happened?	Alcohol	Treatment—promotes behavior change through developing a favorable attitude, experiencing positive social influences, and developing high self-efficacy toward the behavior	The I-Change Model	Social norms, perceived pressure, and self-efficacy	Web-based serious game	Cluster RCT[a] (pre- and posttest measurements)—34 schools with 2649 adolescents (aged between 15 and 19 years) divided into experimental (1622) and control (1027) groups	Not specified	The serious game was effective in reducing binge drinking in adolescents aged 15 and 16 years when they participated in at least two intervention sessions. Interaction effects were found between excessive drinking and educational level and between weekly consumption and age. Additional analyses revealed that prolonged use of the intervention was associated with stronger effects for binge drinking. However, overall adherence to the intervention was low.
Epstein et al [32]—Bacon Brains	Alcohol and drugs	Prevention—education on the science of addiction and how alcohol and other drugs affect the brain	Not specified	Not specified	Web-based serious game	RCT (pre- and posttest measurements) with 244 students (sixth to eighth grade) aged 11 to 15 years	Not specified	A more significant knowledge gain among the intervention groups was found compared with the control group. The intervention was helpful in teaching students core concepts about the science of addiction, and the knowledge they learned persisted through posttest assessment. Girls acquired knowledge gains in the collaborative and competitive game conditions, whereas boys demonstrated similar gains only in the competitive condition. Students in the experimental conditions reported enjoying playing the games more than students in the control condition.
Hookham et al [26,27]—Shadow	Alcohol (comorbidity with depression)	Treatment—promotion of behavioral and cognitive change related to mood and misuse of alcohol	Not specified	Cognitive behavioral therapy and motivational interviewing	Web-based serious game where the storyline is based on branching of predefined dialogues	Comparative study (posttest measurements, within-subject) between a gamified and nongamified version of a web-based alcohol abuse and depression treatment with 10 university students (aged 18-30 years)	Not specified	No significant differences were found between the gamified and nongamified versions of the web-based treatment in terms of usability, ease of use, perceived usefulness, or engagement.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Hughes et al [36]—CollegeLiVE	Alcohol and drugs	Prevention—presents typical situations to practice social skills, protective behaviors, and self-reflection	Social cognitive theory and interactive performance theory	Not specified	Virtual reality scenario based on a computational framework that facilitates the creation and remote control of avatars for interaction purposes	Quasi-experimental study (pre- and posttest measurements) with 68 university students (45 in the intervention group and 23 in the control group) aged ≥18 years	Not specified	No significant differences were found between the intervention and control groups when questioned about specific protective behaviors. Nevertheless, the participants in the intervention group were significantly more likely to implement protective behaviors such as trying to stop somebody they knew from deciding to drive after drinking and using a designated driver than the participants in the control group.
Gamberini et al [50]—no name provided	Alcohol and drugs	Prevention—education on and awareness of the risks related to the consumption of psychoactive substances for partygoers	Not specified	Not specified	Multiplayer web-based serious game	Two pilot studies (pre- and posttest measurements), one to assess risk awareness and user experience with 227 participants (mean age 24.7, SD 6.78 years) and the other to assess knowledge increase with 81 participants (mean age 23.4, SD 2.02 years)	Data collected anonymously	The results from the first pilot study indicate that the user experience was high. In the second pilot study, significant differences were found in risk assessment and knowledge of substance consumption risks and coping skills between pre- and postgame sessions.
Boendermaker et al [49]—Shots	Alcohol	Treatment—cognitive retraining of selective attention toward alcohol	Not specified	CBM-A[b]	Not specified	RCT (pre- and posttest measurements) with 96 heavy-drinking university students (aged 18-28 years). A total of 33 participants were assigned to the gamified version of CBM-A, 30 were assigned to a regular version of CBM-A, and 33 were assigned to a placebo version.	Not specified	There was an overall decline in alcohol attentional bias, but this effect was primarily driven by the regular training condition. Motivation to train decreased equally in all conditions, indicating that the motivational elements in the gamified version could not sufficiently counteract the tiresome nature of the training. Moreover, motivation to change with respect to planning to drink less in the future increased in the regular and placebo training but decreased in the gamified training condition, which may indicate potential detrimental effects of disappointing gamification.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Boyle et al [51]—CampusGANDR	Alcohol	Treatment—modification of behavior by correcting normative perceptions and inducing reductions in alcohol use	Not specified	Personalized normative feedback	Web-based serious game with Facebook log-in credentials to simulate a social game (but no real social connection)	RCT (pre- and posttest measurements and follow-up) with 237 undergraduate students aged between 18 and 24 years. A total of 113 students were assigned to the use of the serious game, and 124 were assigned to the standard intervention (control group).	Not specified	Participants in the serious game condition reported significantly reduced peer drinking norms and alcohol consumption at the 2-week follow-up compared with participants who received the standard intervention. A mediation model demonstrated that this effect was driven by larger reductions in perceived drinking norms among participants assigned to the serious game.
Damasceno et al [37]—no name provided	Drugs	Prevention—education on the damage of drug abuse and submission of the player to moral decisions	Social cognitive theory	Not specified	PC-based serious game	Pilot study (posttest measurements) with 69 school students (mean age 13.7 years)	Not specified	On the basis of the collection and analysis of in-game behaviors (statistics such as the time spent, correct and wrong answers, and response time), the authors present a general picture of the differences in game performance among groups of users. The authors conclude that serious games could be a powerful tool for drug abuse prevention as an instructional mechanism and for the identification of risk behaviors.
Stapinski et al [52]—Pure Rush	Drugs (focused on cannabis, ecstasy, methamphetamine, and hallucinogens)	Prevention—education on the effects associated with each drug and pairing drug-related cues with negative stimuli	Not specified	Not specified	Web-based serious game with a mobile version for Android devices	Feasibility study (posttest measurements) with 25 students aged between 14 and 17 years and an RCT (pre- and posttest measurements) in 9 schools with 281 students aged between 13 and 16 years. A total of 148 students were assigned to game sessions, and 133 were assigned to the control group.	Data collected anonymously	Results from the feasibility study indicate that students enjoyed playing Pure Rush and found the game age-appropriate and the game's infographics understandable and appealing. Results from the RCT indicate significant knowledge increase from before to after the intervention in both groups (no significant difference between groups). Knowledge gain was greater in the game condition than in the control group but only in female students. Very low intentions to use illicit drugs were observed in both groups. No significant differences were found in lesson engagement and in future intentions to use drugs between the intervention and control groups.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Earle et al [42]—CampusGAN-DRV2	Alcohol	Treatment—modification of behavior by correcting normative perceptions and inducing reductions in alcohol use	Self-determination theory	Personalized normative feedback	Mobile-based serious game	RCT (pre- and posttest measurements and follow-up) with 276 first-year university students. A total of 93 students were assigned to the game session receiving standard feedback, 90 were assigned to the game plus supplemented feedback on their perceptions and behaviors, and 93 were assigned to a control group.	Not specified	Participants who were assigned to the game with the supplemented feedback reduced their drinking significantly during the 2 months after the intervention in comparison with control participants. Reduction in drinking behavior was stronger among heavy drinkers.
Duncan et al [53]—SmokeSCREEN	Drugs	Prevention—education and presentation of social situations to develop behavioral skills associated with primary prevention of cigarette and marijuana use	Not specified	Not specified	Mobile-based serious game	Pilot study (pre- and posttest measurements) with 25 adolescents aged between 11 and 14 years	Not specified	Improvements in knowledge were found for both cigarette and marijuana between baseline and after playing. No significant changes in perceived social norms toward both substances were reported. The players provided positive feedback about their experience with the serious game.
Metcalf et al [45]—Take Control	Alcohol and cigarette	Treatment—practice of refusal skills and increase in self-efficacy by denying trigger or cue items in a nonthreatening environment	Not specified	Cue exposure therapy, extinction therapy, virtual reality therapy, and cognitive behavioral therapy	PC-based serious game with Kinect	Quasi-experimental wait-list study (pre- and posttest measurements) with 61 participants aged >18 years. A total of 29 participants were assigned to the intervention group, and 32 were assigned to the wait-list control group.	Data collected anonymously	The results reported that substance use decreased or remained for most users, although more for alcohol consumers than for tobacco users. Participants in recovery for alcohol use reported more benefit than those in recovery for tobacco use, with a statistically significant increase in self-efficacy, attitude, and intended behavior. Participants found the game engaging and fun and felt that playing it would support recovery efforts.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Kapitány-Fövény et al [46]—Once Upon a High	Alcohol and drugs	Prevention—education on the epidemiology and risks of substance use, promotion of health-conscious behavior, and decrease in stigma and negative attitudes and increase in willingness to help peers with substance use problems	Not specified	Self-efficacy	Mobile-based app containing 2 serious games	Quasi-experimental study (pre- and posttest measurements) with 386 students aged between 14 and 18 years from 4 different schools. A total of 255 students were assigned to the intervention group (app use), and 131 were assigned to the control (nonapp use) group.	Data collected anonymously	Users in the intervention group showed a greater decrease in energy drink consumption. No relevant differences between the 2 groups were found in knowledge gaining on psychoactive substances and physical exercise frequency. A correlation between the app's perceived usefulness and a decreasing frequency of past-month alcohol use was found. Results also indicate that a decrease in negative attitudes toward substance users might be a risk factor for increasing past-month alcohol consumption.
Gamberini et al [54]—no name provided	Alcohol and drugs	Prevention—education on the potential risks of psychoactive substance abuse during nightlife events	Not specified	Not specified	Multiplayer web-based serious game	Quasi-experimental comparative study (posttest measurements) with 136 young adults (mean age 23.5 years). A total of 67 participants used the game, and 69 read leaflets with information on the potential risks of psychoactive substance abuse.	Data collected anonymously	The evaluations of the game credibility and effectiveness were positive. No significant difference was found between game and leaflets except for informativeness, on which leaflets performed better than the game.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Abroms et al [48]—Recovery Warrior 2.0	Drugs	Treatment—helping the patients in the development of negative associations with drugs and acquisition of drug refusal skills	Social cognitive theory and reinforcement theory of motivation	Repetition priming	PC-based serious game with Kinect for body motion and voice recognition	RCT (pre- and posttest measurements and follow-up) with 80 participants aged between 15 and 25 years under a drug treatment program. A total of 36 participants were assigned to the use of the game+treatment as usual, and 44 were assigned to the control group (treatment as usual only).	Not specified	Participants in the intervention group mostly agreed that they would use the refusal skills taught by the game and reported attending more outpatient counseling sessions than those in the control group, but the difference was not significant. Cravings declined for both groups from baseline to the 4- and 8-week follow-up, but the differences between groups were not statistically significant. Self-efficacy fluctuated slightly but did not change widely between the intervention and control groups. The game had no effect on drug use at 4 or 8 weeks after discharge, with the exception of a benefit reported at the 4-week follow-up among participants receiving treatment for marijuana addiction.
Carvalho et al [38]—JiB	Alcohol	Prevention—education on the consequences of alcohol abuse promoting empowerment of the individual, traditional cultures, and social responsibility	Not specified	Not specified	Mobile-based serious game	Pilot study (pre- and posttest measurements) with 23 participants aged between 20 and 29 years	Data collected anonymously	The results were analyzed by dividing the participants into habitual and nonhabitual players. The game presented more positive than negative effects on all users. For habitual players, the game did not have a high level of challenge owing to the low difficulty and small learning curve. A higher game tension was found in nonhabitual players versus habitual players.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Willmott et al [47]—Perfect Pour and Dumb Driver	Alcohol	Prevention—education on the physiological effects of varying levels of blood alcohol concentration on driving ability (Dumb Driver) and on the standard alcohol content of 6 types of alcoholic drinks through a multilevel test of pouring accuracy (Perfect Pour)	Theory of reasoned action	Subjective norms and attitudes	Web-based serious game	Pilot study (pre- and posttest measurements) with 303 students aged between 12 and 17 years	Not specified	A positive relationship between average game duration and attitude was found, indicating that the longer the participants played Perfect Pour, the more positive their attitudes became toward binge drinking. On the contrary, a negative relationship between average score and attitudes indicates that the higher the players scored in Perfect Pour, the more negative their attitudes became toward binge drinking. No significant associations were observed among gameplay metrics, attitudes, and subjective norms for Dumb Driver.
Tan et al [55]—Drug Defense	Alcohol	Prevention—education on the consequences of excessive alcohol use	Social cognitive theory and multiple intelligences theory	Not specified	Mobile-based serious game	Pilot test (pre- and posttest measurements) with 69 university students (aged 18 to 21 years)	Not mentioned	Significant differences between pre- and posttest measurements were found, indicating an increase in knowledge of alcohol. Playability was good according to the values obtained in the evaluation of gameplay, story, and mechanics of the game.
Mostajeran et al [40]—no name provided	Alcohol	Treatment—minimize the occurrence of relapse by practicing to avoid alcohol in a simulated supermarket	Self-determination theory	Cue exposure therapy and approach-avoidance training	PC-based serious game with virtual reality using a head-mounted display	Comparative study (posttest measurements, within-subject) with 13 participants (aged 22 to 35 years). All the participants were assigned to a gamified version of the approach-avoidance training and cue exposure therapy and a nongamified version of the approach-avoidance training.	Not mentioned	All the gamified versions received high usability scores. The gamified version of the approach-avoidance training was more cognitively demanding than the nongamified version. The gamified version of the approach-avoidance training was more motivating than the nongamified version. Participants made fewer errors in the gamified version of the approach-avoidance training than in the nongamified version. The users preferred the mini-game of the approach-avoidance training than the mini-game of the cue exposure therapy.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Yap et al [35]—Drug Defense	Alcohol	Prevention—education on the consequences of excessive alcohol use	Social cognitive theory and multiple intelligences theory	Not specified	Mobile-based serious game	RCT (pre- and posttest measurements) with 140 university students (aged 18 to 21 years). A total of 69 were assigned to the use of the mobile-based game, and 71 were assigned to a video intervention (control group).	Encoding of data and password protection	Participants who used the game showed a significant increase in knowledge scores and a decrease in intent to use but not in actual use. Participants in the video intervention reported a significant increase in knowledge and a decrease in both intent and use. Findings also showed a significant difference in alcohol knowledge for both game and video settings, with the game having a larger effect size than the video.
Ozer et al [33]—INSPIRE	Alcohol	Prevention—education on personal efficacy and skills by providing opportunities to practice strategies for reducing risky health behaviors	Social cognitive theory	Self-efficacy and self-regulation	PC-based serious game with interactive narrative experiences	Pilot study (posttest measurements) with 20 adolescents (aged 14 to 19 years)	Not specified	Trace-log data and self-report questionnaires indicate that participants found the game to be engaging, believable, and relevant to their lives. Within the game, participants also successfully used a range of strategies for avoiding alcohol use. The participants accessed fewer in-game objects than anticipated and spent less time examining the associated infographics designed to affect knowledge of the effects of alcohol use.
Hong et al [34]—One Shot	Alcohol	Prevention—education on and practice of drinking refusal self-efficacy	Social cognitive theory and protection motivation theory	Not specified	Web-based serious game	Quasi-experimental study (1-group, pre- and posttest measurements) with 550 young adults (aged 21 to 25 years) at risk of binge drinking	Not specified	Results show improvements from before to after the game in intention to drink less and in 5 of the 7 indicators of drinking refusal self-efficacy. Risky alcohol decisions within the game and game time predicted enjoyment, which, in turn, predicted intention to drink less and drinking refusal self-efficacy. Enjoyment significantly mediated the effects of game time and risky alcohol decisions on intention to drink less and drinking refusal self-efficacy.

XSL•FO
RenderX

Study and game name	Substance addressed	Type of intervention	Theoretical basis		Computational techniques	Evaluation protocol	Mechanism for data security and confidentiality	Main reported results
			Theory	Theoretical construct or intervention technique				
Boyle et al [43]—GANDR	Alcohol	Treatment—modification of behavior by correcting normative perceptions and inducing reductions in alcohol use	Self-determination theory	Personalized normative feedback	Web application (multiplatform) serious game	RCT (pretest measurements and follow-up 1 month later) with 223 first-year university alcohol-experienced students (mean age 18.05 years). A total of 74 students were assigned to the gamified-only intervention, 74 were assigned to the gamified social media intervention, and 75 were assigned to the nongamified intervention (control group).	Not specified	Reported results indicate a significantly greater reduction in drinking at follow-up in the gamified social media condition than in the traditional (nongamified) condition. It was unclear whether the gamified-only condition would lead to greater reductions in drinking than the traditional condition. The gamified-only condition did not lead to a significantly greater reduction in drinking at follow-up compared with the nongamified condition. Relative to the nongamified condition, the gamified social media condition led to significantly greater reductions in drinking for those who were lighter drinkers with less exposure to alcohol-related content on social media.
Boyle et al [44]—LezParlay	Alcohol	Treatment—reduction of risks related to alcohol consumption	Self-determination theory	Personalized normative feedback	Web application (multiplatform) serious game	RCT (pretest measurements and follow-up at 2 and 4 months) with 499 LBQ[c] women aged 21 to >40 years with moderate and heavy alcohol consumption. A total of 143 women were assigned to the gamified intervention on alcohol use and stigma coping, 179 were assigned to the gamified intervention on alcohol use only, and 177 were assigned to the control group.	Not specified	Obtained results indicate that participants who received the intervention on alcohol use and both alcohol use and stigma coping had similar reductions in their weekly drinks, peak drinks, and negative consequences relative to those in the control group at the 2-month follow-up. However, at the 4-month follow-up, reductions in alcohol consumption outcomes faded among those who received the alcohol-only intervention, whereas they remained relatively robust among those who received both the alcohol use and coping intervention. Regarding feasibility, the participants reported the competition to be highly acceptable and psychologically beneficial as a whole.

[a]RCT: randomized controlled trial.

[b]CBM-A: cognitive bias modification of attention.

[c]LBQ: lesbian, bisexual, and queer.

XSL•FO
RenderX

Computer-Based Techniques Used

In most analyzed studies (24/29, 83%), little technical detail was provided about the computer-based methods or techniques used to develop the serious games. Most of the studies only mentioned the platform where the serious game could be executed: web-based (11/29, 38%), mobile app (7/29, 24%), multi-platform (3/29, 10%), or PC-based (2/29, 7%). The studies that provided more technical detail mentioned the use of virtual reality for the creation of immersive environments [31,40], the use of a Kinect device for a motion-based interaction [45,48], the implementation of branching predefined dialogues to create interactive narrative experiences [26,27,33], the integration with Facebook to simulate a social serious game [51], and the use of a computational framework [56] that facilitated the creation and remote control (as in Wizard-of-Oz scenarios) of avatars for interaction purposes [36].

Data Security and Confidentiality

Most of the reviewed studies (19/29, 66%) did not explicitly mention whether a specific mechanism was implemented in the serious game to assure the data protection and confidentiality of the users. Although all the studies (29/29, 100%) reported that an institutional ethical committee had approved the protocol implemented to evaluate the serious game, only some (9/29, 31%) mentioned the type of data collected from the participants and how confidentiality was guaranteed. The most frequent type of data collected in these studies was sociodemographic information (eg, sex, age, and ethnicity) without collecting identifiable information, maintaining the anonymity of the participants [29,30,38,39,45,46,50,52,54]. Only 3% (1/29) of the studies mentioned encoding of the data and password protection [35].

Evaluation Procedures and Reported Results

The analyzed studies reported the evaluation of the serious games through RCTs (10/29, 34%), quasi-experimental studies (5/29, 17%), pilot studies (11/29, 38%), comparative studies (2/29, 7%), and a proof-of-concept study (1/29, 3%). The reported results from the RCTs were in general positive, describing significant differences between the intervention and control groups in terms of decreasing binge drinking [41], reducing peer drinking norms and alcohol use [42-44,51], gaining knowledge on how alcohol and drugs affect the brain [32], increasing substance-related knowledge (but not decreasing psychoactive substance use) [46], being more accepting toward using refusal skills of drug use (but no effect was found on drug use) [48], or increasing knowledge of the effects of alcohol use and protective behavioral techniques and decreasing intention of alcohol use (but not actual use) [35].

Nevertheless, some of these RCT studies reported mixed results in the use of serious games: the longer the average game duration, the more positive the attitude toward binge drinking that was found. However, the same study reported that the higher the players scored in the game, the more negative their attitudes became toward binge drinking [47]. Similarly, no significant differences were found in learning outcomes on drug education (except for women, where knowledge gaining was greater in the intervention than in the control group), lesson engagement, or future intentions to use drugs between the intervention and control groups [52]. Moreover, the results reported in the study by Boendermaker et al [49] indicate that a gamified version of a cognitive bias modification of attention program reduced the motivation to train, and self-reported drinking behavior was not affected.

In addition, the results reported from quasi-experimental studies reflected positive outcomes after the use of serious games. The participants reported in the study by Klisch et al [29] demonstrated an increase in negative attitudes toward prescription drug abuse after the use of serious games, and the games' ratings for satisfaction and engagement were above average. Alcohol use decreased in individuals under substance use treatment during an intervention using a serious game, and there was a significant increase in self-efficacy, attitude, and behavior to prevent relapse. Nevertheless, these positive results decreased over time after the last playing session [45]. The study by Hong et al [34] reported that participants improved from before to after the game in intention to drink less alcohol and in drinking refusal self-efficacy, where the enjoyment of the game was a key feature. The results reported in the study by Hughes et al [36] indicate that participants from the intervention group were more likely to adopt protective behaviors such as trying to stop somebody from driving after drinking and using a designated driver than participants in the control condition. One of these studies also demonstrated that participants in a game session to provide information about potential risks of substance abuse during nightlife events evaluated the serious game to be as positive and credible as more traditional information tools such as leaflets [54].

The rest of the papers (11/29, 38%) reported pilot studies where the developed serious games were assessed with a group of participants in terms of usability and acceptability, including outcomes such as perceived utility, engagement, enjoyment, or game experience [33,38,39]; the performance of the players to assess the suitability of serious games as a learning tool for the prevention of drug consumption [37]; and the effectivity in terms of change in knowledge and attitudes toward substance use [30,31,50,53]. Finally, the 7% (2/29) of comparative studies assessed the differences between a gamified and a nongamified version of existent treatment programs. One of these studies did not find relevant differences in usability, ease of use, and perceived usefulness between the 2 treatment versions [26,27]. The other comparative study reported differences in motivation, cognitive demand, and the number of errors made between the 2 versions. The gamified version was more motivating and cognitively demanding, and the participants made fewer mistakes [40]. Table 2 summarizes the main extracted data from each analyzed study.

Discussion

Principal Findings

Given the increasing interest in developing serious games as digital interventions to support the prevention and treatment of alcohol and drug consumption, it is relevant for future works in the area to identify key features that designers of serious games should consider. This scoping review contributes by

XSL•FO
RenderX

identifying and summarizing the description of 7 main characteristics that reflect how these games are currently designed and what are the intervention results reported. The increase in studies focused on developing and evaluating this type of serious game is evident in this review. Previous similar reviews (although considering only serious games with learning purposes) published in 2014 [18] and 2016 [19] found 12 and 8 serious games, respectively.

By contrast, our review found 29 serious games considering only the previous 12 years (2010-2022), with most (24/29, 83%) published in the last 6 years, from 2016 onward. The main substance addressed in the serious games included in this review was alcohol by itself or with other drugs. This result is not surprising as alcohol, a licit substance, is still the most consumed by the young population worldwide, and it causes more direct and indirect deaths than illicit drugs [57]. As alcohol and drugs share many risk behaviors and health consequences, serious game scenarios and narratives can be applied to both types of substances. Nevertheless, it is necessary to carry out more specific studies to assess if any positive effects from serious games could be applied equally to alcohol and illicit drugs without changing much of the game's mechanics.

Two-thirds of the 29 analyzed serious games (18/29, 62%) focused on supporting preventive interventions, whereas the rest (11/29, 38%) were developed to complement the treatment process. This disparity could be because preventive serious games are mainly developed to support educative aspects associated with, for example, risk behaviors, protective skills, and the negative consequences of alcohol and drug abuse. As learning material, these serious games do not necessarily require the direct or close supervision of a specialist and have the potential to reach a high number of the intended population (as part of formal curricula or not). By contrast, the serious games focused on treatment must be carefully designed considering the characteristics of potential users with high consumption levels or addiction problems. Moreover, to maximize the effectiveness and security of using these serious games, the treatment provider must necessarily be involved in the stages of the treatment addressed by the game. In this sense, serious games are not as different from other digital-based tools where their benefits to the treatment process are higher if used under specialist guidance or supervision [58,59].

Using a background theory in any intervention, including serious game–based interventions, is essential to understand not only what interventions work but why they work [60]. In this sense, it is noticeable that most of the reviewed studies (22/29, 76%) mentioned the theoretical basis (theory, theoretical constructs, or intervention techniques or all three) used in the serious games. The theoretical background of the reviewed serious games also depended on their prevention or treatment support objectives. For the games used with prevention objectives, different theories were used to design the games' mechanics and contents, highlighting the SCT. The main posits of the SCT (formerly known as social learning theory) consider that the acquisition and maintenance of specific (healthy or not healthy) behaviors depend on reciprocal interaction between the individual and their social context mediated by positive and negative reinforcements [61]. Thus, this theory offers promising opportunities to represent this interaction through different game scenes where the player can observe (and learn) the consequences of their own or others' actions (vicarious learning) in realistic scenarios associated with substance use.

It is noteworthy that 39% (7/18) of the analyzed serious games with prevention objectives (through education) did not explicitly mention any theoretical background. This lack of specification would make it difficult to further analyze what specific game mechanics (and the theoretical tenets on which they are based) positively affect the learning process to prevent substance use. By contrast, the serious games developed to support the treatment process (11/29, 38%) were based on different intervention techniques, with none standing out from the others. Depending on the desired outcomes, different theoretical constructs and intervention techniques were used, including cognitive behavioral therapy and motivational interviewing to promote change of behaviors related to substance abuse or cognitive bias modification for cognitive retraining of selective attention toward substances, to name a few. Some others (5/29, 17%) took advantage of the user interface's characteristics to implement, for example, cue exposure therapy and approach-avoidance training in virtual reality scenarios accessed through Kinect or head-mounted displays.

Regarding the computational techniques and approaches used, most of the included serious games (23/29, 79%) only specified the platform where the game could be deployed: web, mobile, or stand-alone environments. Some of them also specified the implementation of multiplayer or social network characteristics and virtual reality scenarios. Nevertheless, none of the reviewed serious games implemented more advanced techniques and interfaces such as brain-computer interaction and eye and head tracking combined with machine learning algorithms to, for example, automatically detect individual characteristics (eg, mood, emotions, and attention) of the players and personalize or adapt the game contents and mechanics. These techniques are currently used in serious games with evident advantages in other contexts, particularly for learning purposes [62-64]. Thus, a future direction would be to develop and evaluate serious games implementing these computational approaches and assess whether their use improves the desired outcomes in preventing and treating substance use.

The use of serious games for health raises particular challenges as their objective is not only to entertain but also to influence users' attitudes and behaviors that would affect their everyday life [65]. One of these challenges is to decide what data are collected through the serious game, how this information is used, and how to guarantee user data security and confidentiality. It is noticeable that only 34% (10/29) of the included studies mentioned issues related to privacy. This result, of course, does not mean that the rest of the reviewed serious games did not implement a data security and confidential mechanism as all the reviewed works (29/29, 100%) reported the approval of their use in studies involving humans by an ethical board. A protocol for ethical approval in digital interventions usually must define what data are collected, where they are stored, who will use them, and how privacy and confidentiality are guaranteed [66]. Nevertheless, it seems necessary that this information be included when reporting the

XSL•FO
RenderX

design, implementation, and evaluation of serious games for health to identify how the challenge of data security and privacy is addressed instead of only mentioning that the data are collected anonymously.

Another of the relevant current and future challenges in developing these types of serious games is to evaluate their actual efficacy on the addressed population. The assessment of evaluation results is particularly relevant for the effective adoption and appropriation of serious games (as digital technology for health), where evidence of the contributions of digital health interventions to the performance of health systems and their impact on people's health and well-being is one of the key activities of the WHO global strategy on digital health [67]. Many of the reviewed papers (11/29, 38%) reported feasibility and pilot studies focused on, for example, usability, engagement, game experience, or perceived utility of the serious games. There were a few quasi-experimental studies (5/29, 17%) where the main factors assessed were changes in intentions or attitudes toward substance consumption and knowledge gaining. Overall, the reported results from these quasi-experimental studies were positive considering the observed differences in the participants at pre- and posttest measurements. Still, all these studies warrant a follow-up assessment to know whether these observed effects are maintained in the long run.

Similarly, the reported RCTs also highlighted some positive effects when comparing the intervention and control groups in terms of reducing substance use, increasing knowledge, or being more accepting of refusal skills. Nevertheless, other RCTs (3/10, 30%) did not find significant differences, and 10% (1/10) even warned about gamification's detrimental effects [49]. A follow-up assessment was included in 50% (5/10) of the reported RCTs, assessing the long-term effect after 2 months of the intervention in 60% (3/5) of them. One of these studies reported that cravings declined from baseline to the 4- and 8-week follow-up, although differences between the intervention and control groups were not significant [48]. Another study reported that, after 4 months, reductions in alcohol consumption faded in the group of participants that received the alcohol-only intervention and remained *relatively* robust in the group that received the alcohol and coping stigma intervention [44]. Thus, more RCT studies are necessary focusing on identifying relevant outcomes such as changes in attitudes, motivation, and behaviors. Moreover, including follow-up assessments for more prolonged periods and the reproducibility of these studies would help understand better the effect of serious game–based interventions for both prevention and treatment of substance use.

Limitations

This study has some limitations related to the selection procedure. One of these limitations is that only studies reporting an evaluation procedure were included. Thus, some serious games that implemented more sophisticated computational techniques, such as machine learning, other artificial intelligence methods, or other interactive devices that had not yet been evaluated, were not considered. Moreover, the only sources used to identify serious games were academic repositories. No other sources such as app stores were considered to assess whether some of these available apps included gamification techniques aimed at preventing or supporting substance use treatment. Nevertheless, the contribution of this study is relevant as it highlights the current state of the art in the research and development of this type of serious games, identifying relevant characteristics such as the substance addressed, their theoretical roots, the computational techniques used, the mechanism for data security and confidentiality, and the results obtained from an evaluation protocol.

Conclusions

The development and evaluation of serious games to support prevention and treatment interventions for substance use have increased in the last decade. The review presented in this paper describes a set of main characteristics of 29 serious games evaluated with adolescents and young adults aiming to prevent and reduce the consumption of alcohol and drugs. Most of the analyzed serious games were designed to prevent or reduce alcohol consumption. Those developed with prevention objectives were also the majority compared with those developed to support treatment. The reported evaluation protocols included mostly pilot studies, quasi-experimental studies, and RCTs.

Overall, most of the reviewed serious games reported positive results regarding acceptability, usability, increased knowledge, and change in attitudes and behaviors toward alcohol and drug consumption. An area of future research is the incorporation of other human-computer interaction and artificial intelligence techniques to identify relevant user data, allowing for the personalization of the offered interventions. In addition, more studies—that facilitate reproducibility—are necessary to better identify the long-term effects of these serious games (and whether boosted playing sessions are necessary after some time) and what specific game mechanics are more useful for preventive or treatment interventions.

Acknowledgments

This work was funded by the National Council of Science and Technology of Mexico through the Strategic National Programs-Health under research project 3210 *Desarrollo y evaluación de una plataforma tecnológica de ayuda a la detección, seguimiento e intervención temprana de problemas de salud mental y adicciones en la comunidad escolar, primer y segundo nivel de atención* ("Development and evaluation of a technological platform to support the detection, follow-up, and early intervention of mental health and addiction problems in the school community, primary and specialized health units").

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. World Drug Report 2021. United Nations. URL: https://www.unodc.org/unodc/en/data-and-analysis/wdr2021.html [accessed 2022-08-10]
2. Poznyak V, Rekve D. Global status report on alcohol and health 2018. World Health Organization. 2018 Sep 27. URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2022-08-09]
3. Winters KC, Arria A. Adolescent brain development and drugs. Prev Res 2011;18(2):21-24 [FREE Full text] [Medline: 22822298]
4. Casey BJ, Jones RM, Hare TA. The adolescent brain. Ann N Y Acad Sci 2008 Mar;1124:111-126 [FREE Full text] [doi: 10.1196/annals.1440.010] [Medline: 18400927]
5. Adolescent and young adult health. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2022-08-09]
6. Curtin SC, Tejada-Vera B, Warmer M. Drug overdose deaths among adolescents aged 15-19 in the United States: 1999-2015. NCHS Data Brief 2017 Aug(282):1-8 [FREE Full text] [Medline: 29155681]
7. Behrendt S, Wittchen HU, Höfler M, Lieb R, Beesdo K. Transitions from first substance use to substance use disorders in adolescence: is early onset associated with a rapid escalation? Drug Alcohol Depend 2009 Jan 01;99(1-3):68-78. [doi: 10.1016/j.drugalcdep.2008.06.014] [Medline: 18768267]
8. Marsch LA, Borodovsky JT. Technology-based interventions for preventing and treating substance use among youth. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):755-768 [FREE Full text] [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
9. Belendiuk KA, Riggs P. Treatment of adolescent substance use disorders. Curr Treat Options Psychiatry 2014 Jun 01;1(2):175-188 [FREE Full text] [doi: 10.1007/s40501-014-0016-3] [Medline: 24855595]
10. Hopson L, Wodarski J, Tang N. The effectiveness of electronic approaches to substance abuse prevention for adolescents. J Evid Inf Soc Work 2015;12(3):310-322. [doi: 10.1080/15433714.2013.857178] [Medline: 25661894]
11. Bishop D, Bryant KS, Giles SM, Hansen WB, Dusenbury L. Simplifying the delivery of a prevention program with Web-based enhancements. J Prim Prev 2006 Jul;27(4):433-444. [doi: 10.1007/s10935-006-0042-z] [Medline: 16763766]
12. Bandawar M, Narasimha VL, Chand P. Use of digital technology in addiction disorders. Indian J Psychiatry 2018 Feb;60(Suppl 4):S534-S540 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_21_18] [Medline: 29540927]
13. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertain Comput 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
14. Espinosa-Curiel IE, Pozas-Bogarin EE, Lozano-Salas JL, Martínez-Miranda J, Delgado-Pérez EE, Estrada-Zamarron LS. Nutritional education and promotion of healthy eating behaviors among Mexican children through video games: design and pilot test of FoodRateMaster. JMIR Serious Games 2020 Apr 13;8(2):e16431 [FREE Full text] [doi: 10.2196/16431] [Medline: 32281539]
15. Holtz BE, Murray K, Park T. Serious games for children with chronic diseases: a systematic review. Games Health J 2018 Oct;7(5):291-301. [doi: 10.1089/g4h.2018.0024] [Medline: 29957082]
16. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 01;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
17. Gaspar JD, Lage EM, Silva FJ, Mineiro É, Oliveira IJ, Oliveira I, et al. A mobile serious game about the pandemic (COVID-19 - did you know?): design and evaluation study. JMIR Serious Games 2020 Dec 22;8(4):e25226 [FREE Full text] [doi: 10.2196/25226] [Medline: 33301416]
18. Rodriguez DM, Teesson M, Newton NC. A systematic review of computerised serious educational games about alcohol and other drugs for adolescents. Drug Alcohol Rev 2014 Mar;33(2):129-135. [doi: 10.1111/dar.12102] [Medline: 24329810]
19. Pasquim HM, Soares CB, Santoro R. Drug education games for youth. Soc Med 2016;10(2):61-69.
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Drugs (psychoactive). World Health Organization. URL: https://www.who.int/health-topics/drugs-psychoactive [accessed 2022-08-09]
22. Society for Adolescent Health and Medicine. Young adult health and well-being: a position statement of the Society for Adolescent Health and Medicine. J Adolesc Health 2017 Jun;60(6):758-759. [doi: 10.1016/j.jadohealth.2017.03.021] [Medline: 28532650]
23. Fomiatti R, Farrugia A, Fraser S, Hocking S. Improving the effectiveness and inclusiveness of alcohol and other drug outreach models for young people: a literature review. Drugs (Abingdon Engl) (forthcoming) 2021 Sep 08:1-10. [doi: 10.1080/09687637.2021.1975652]

XSL•FO
RenderX

24. Hendriks H, Gebhardt WA, van den Putte B. Alcohol-related posts from young people on social networking sites: content and motivations. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):428-435. [doi: 10.1089/cyber.2016.0640] [Medline: 28650223]
25. Poelen EA, Scholte RH, Engels RC, Boomsma DI, Willemsen G. Prevalence and trends of alcohol use and misuse among adolescents and young adults in the Netherlands from 1993 to 2000. Drug Alcohol Depend 2005 Sep 01;79(3):413-421. [doi: 10.1016/j.drugalcdep.2005.03.020] [Medline: 16039803]
26. Hookham G, Nesbitt K, Kay-Lambkin F. Comparing usability and engagement between a serious game and a traditional online program. In: Proceedings of the Australasian Computer Science Week Multiconference. 2016 Presented at: ACSW '16; February 1-5, 2016; Canberra, Australia p. 1-10. [doi: 10.1145/2843043.2843365]
27. Hookham G, Kay-Lambkin F, Blackmore K, Nesbitt K. Using startle probe to compare affect and engagement between a serious game and an online intervention program. In: Proceedings of the Australasian Computer Science Week Multiconference. 2016 Presented at: ACSW '16; February 1-5, 2016; Canberra, Australia p. 1-9. [doi: 10.1145/2843043.2843481]
28. Elias-Lambert N, Boyas JF, Black BM, Schoech RJ. Preventing substance abuse and relationship violence: proof-of-concept evaluation of a social, multi-user, tablet-based game. Child Youth Serv Rev 2015 Jun;53:201-210. [doi: 10.1016/j.childyouth.2015.04.007]
29. Klisch Y, Bowling KG, Miller LM, Ramos MA. The impact of science education games on prescription drug abuse attitudes among teens: a case study. J Drug Educ 2013;43(3):255-275. [doi: 10.2190/DE.43.3.d] [Medline: 25445507]
30. Klisch Y, Miller LM, Wang S, Epstein J. The impact of a science education game on students' learning and perception of inhalants as body pollutants. J Sci Educ Technol 2012 Apr 01;21(2):295-303 [FREE Full text] [doi: 10.1007/s10956-011-9319-y] [Medline: 23926416]
31. Cheng MT, Annetta L, Folta E, Holmes SY. Drugs and the brain: learning the impact of methamphetamine abuse on the brain through a virtual brain exhibit in the museum. Int J Sci Educ 2010 Jul 16;33(2):299-319. [doi: 10.1080/09500693.2010.482571]
32. Epstein J, Noel J, Finnegan M, Watkins K. Bacon brains: video games for teaching the science of addiction. J Child Adolesc Subst Abuse 2016;25(6):504-515 [FREE Full text] [doi: 10.1080/1067828X.2015.1103348] [Medline: 28603405]
33. Ozer EM, Rowe J, Tebb KP, Berna M, Penilla C, Giovanelli A, et al. Fostering engagement in health behavior change: iterative development of an interactive narrative environment to enhance adolescent preventive health services. J Adolesc Health 2020 Aug;67(2S):S34-S44 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.04.022] [Medline: 32718513]
34. Hong T, Cabrera J, Beaudoin CE. The role of enjoyment in a serious game for binge drinking prevention: pretest-posttest quasi-experimental study. J Med Internet Res 2020 Nov 30;22(11):e21652 [FREE Full text] [doi: 10.2196/21652] [Medline: 33252348]
35. Yap AG, Roy RE, Lasala JR, Tan DK, Hechanova MR, Diy WD, et al. Evaluation of a cognitive-behavioral game design-based mobile game on alcohol use for adolescents. Games Health J 2020 Oct;9(5):353-357. [doi: 10.1089/g4h.2019.0091] [Medline: 33054488]
36. Hughes CE, Hall T, Ingraham K, Epstein JA, Hughes DE. Enhancing protective role-playing behaviors through avatar-based scenarios. In: Proceedings of the 2016 IEEE International Conference on Serious Games and Applications for Health. 2016 Presented at: SeGAH '16; May 11-13, 2016; Orlando, FL, USA p. 1-6. [doi: 10.1109/segah.2016.7586229]
37. Damasceno EF, Nardi PA, Silva AC, L'Erario A. A serious game to support the drug misuse prevention for teenagers students. In: Proceedings of the 2017 IEEE Frontiers in Education Conference. 2017 Presented at: FIE '17; October 18-21, 2017; Indianapolis, IN, USA p. 1-6. [doi: 10.1109/fie.2017.8190441]
38. Carvalho DB, Domingueti DB, Almeida Santos SM, Dias DR. A game to deal with alcohol abuse (Jib): development and game experience evaluation. JMIR Serious Games 2019 Oct 15;7(4):e11151 [FREE Full text] [doi: 10.2196/11151] [Medline: 31617849]
39. Sanchez RP, Bartel CM. The feasibility and acceptability of "Arise": an online substance abuse relapse prevention program. Games Health J 2015 Apr;4(2):136-144 [FREE Full text] [doi: 10.1089/g4h.2014.0015] [Medline: 26181807]
40. Mostajeran F, Kirsten A, Steinicke F, Gallinat J, Kühn S. Towards gamified alcohol use disorder therapy in virtual reality: a preliminary usability study. In: Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces. 2019 Presented at: VR '19; March 23-27, 2019; Osaka, Japan p. 1471-1476. [doi: 10.1109/vr.2019.8797817]
41. Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a Web-based computer-tailored game to reduce binge drinking among Dutch adolescents: a cluster randomized controlled trial. J Med Internet Res 2016 Feb 03;18(2):e29 [FREE Full text] [doi: 10.2196/jmir.4708] [Medline: 26842694]
42. Earle AM, LaBrie JW, Boyle SC, Smith D. In pursuit of a self-sustaining college alcohol intervention: deploying gamified PNF in the real world. Addict Behav 2018 May;80:71-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.005] [Medline: 29407688]
43. Boyle SC, LaBrie JW, Baez S, Taylor JE. Integrating social media inspired features into a personalized normative feedback intervention combats social media-based alcohol influence. Drug Alcohol Depend 2021 Nov 01;228:109007. [doi: 10.1016/j.drugalcdep.2021.109007] [Medline: 34500245]

XSL•FO
RenderX

44. Boyle SC, LaBrie JW, Trager BM, Costine LD. A gamified personalized normative feedback app to reduce drinking among sexual minority women: randomized controlled trial and feasibility study. J Med Internet Res 2022 May 13;24(5):e34853 [FREE Full text] [doi: 10.2196/34853] [Medline: 35559854]
45. Metcalf M, Rossie K, Stokes K, Tallman C, Tanner B. Virtual reality cue refusal video game for alcohol and cigarette recovery support: summative study. JMIR Serious Games 2018 Apr 16;6(2):e7 [FREE Full text] [doi: 10.2196/games.9231] [Medline: 29661748]
46. Kapitány-Fövény M, Vagdalt E, Ruttkay Z, Urbán R, Richman MJ, Demetrovics Z. Potential of an interactive drug prevention mobile phone app (once upon a high): questionnaire study among students. JMIR Serious Games 2018 Dec 04;6(4):e19 [FREE Full text] [doi: 10.2196/games.9944] [Medline: 30514697]
47. Willmott T, Russell-Bennett R, Drennan J, Rundle-Thiele S. The impact of serious educational gameplay on adolescent binge drinking intentions: a theoretically grounded empirical examination. Health Educ Behav 2019 Feb;46(1):114-125. [doi: 10.1177/1090198118780493] [Medline: 30027760]
48. Abroms LC, Fishman M, Vo H, Chiang SC, Somerville V, Rakhmanov L, et al. A motion-activated video game for prevention of substance use disorder relapse in youth: pilot randomized controlled trial. JMIR Serious Games 2019 May 23;7(2):e11716 [FREE Full text] [doi: 10.2196/11716] [Medline: 31124471]
49. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 06;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
50. Gamberini L, Spagnolli A, Nucci M, DeGiuli G, Villa C, Monarca V, et al. A gamified solution to brief interventions for nightlife well-being. In: Proceedings of the 11th International Conference on Persuasive Technology. 2016 Presented at: PERSUASIVE '16; April 5-7, 2016; Salzburg, Austria p. 230-241. [doi: 10.1007/978-3-319-31510-2_20]
51. Boyle SC, Earle AM, LaBrie JW, Smith DJ. PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, Web-based personalized normative feedback alcohol interventions. Addict Behav 2017 Apr;67:8-17 [FREE Full text] [doi: 10.1016/j.addbeh.2016.11.024] [Medline: 27978426]
52. Stapinski LA, Reda B, Newton NC, Lawler S, Rodriguez D, Chapman C, et al. Development and evaluation of 'Pure Rush': an online serious game for drug education. Drug Alcohol Rev 2018 Apr;37 Suppl 1:S420-S428. [doi: 10.1111/dar.12611] [Medline: 28986963]
53. Duncan LR, Hieftje KD, Pendergrass TM, Sawyer BG, Fiellin LE. Preliminary investigation of a videogame prototype for cigarette and marijuana prevention in adolescents. Subst Abus 2018;39(3):275-279 [FREE Full text] [doi: 10.1080/08897077.2018.1437862] [Medline: 29425481]
54. Gamberini L, Nucci M, Zamboni L, DeGiuli G, Cipolletta S, Villa C, et al. Designing and testing credibility: the case of a serious game on nightlife risks. In: Proceedings of the 13th International Conference on Persuasive Technology. 2018 Presented at: PERSUASIVE '18; April 18-19, 2018; Waterloo, ON, Canada p. 213-226. [doi: 10.1007/978-3-319-78978-1_18]
55. Tan DK, Roy RE, Lasala JR, Yap AG, Rodrigo MM, Diy WD, et al. Drug defense: a mobile game for prevention of alcohol abuse. In: Proceedings of the 27th International Conference on Computers in Education Conference. 2019 Presented at: ICCE '19; December 2-6, 2019; Kenting, Taiwan p. 525-530.
56. Nagendran A, Pillat R, Kavanaugh A, Welch G, Hughes C. A unified framework for individualized avatar-based interactions. Presence (Camb) 2014 Aug 01;23(2):109-132. [doi: 10.1162/pres_a_00177]
57. Ritchie H, Roser M. Drug Use. Our World In Data. 2019. URL: https://ourworldindata.org/drug-use [accessed 2022-08-10]
58. Cavanagh K, Herbeck Belnap B, Rothenberger SD, Abebe KZ, Rollman BL. My care manager, my computer therapy and me: the relationship triangle in computerized cognitive behavioural therapy. Internet Interv 2018 Mar;11:11-19 [FREE Full text] [doi: 10.1016/j.invent.2017.10.005] [Medline: 30135755]
59. Kristianslund K. Effective structure and contents of guidance in the treatment of common mental disorders with Internet-based cognitive behavioral therapy: a review of the literature. Universitetet I Oslo. 2019. URL: https://www.duo.uio.no/bitstream/handle/10852/68705/Kristina-Kristianslund-V19.pdf?sequence=1&isAllowed=y [accessed 2022-08-10]
60. Tebb KP, Erenrich RK, Jasik CB, Berna MS, Lester JC, Ozer EM. Use of theory in computer-based interventions to reduce alcohol use among adolescents and young adults: a systematic review. BMC Public Health 2016 Jun 17;16:517 [FREE Full text] [doi: 10.1186/s12889-016-3183-x] [Medline: 27317330]
61. Schunk DH. Social cognitive theory. In: Harris KR, Graham S, Urdan T, McCormick CB, Sinatra GM, Sweller J, editors. APA Educational Psychology Handbook, Vol. 1. Theories, Constructs, and Critical Issues. Washington, DC, USA: American Psychological Association; 2012:101-123.
62. Streicher A, Leidig S, Roller W. Eye-tracking for user attention evaluation in adaptive serious games. In: Proceedings of the 13th European Conference on Lifelong Technology-Enhanced Learning. 2018 Presented at: EC-TEL '18; September 3-5, 2018; Leeds, UK p. 583-586. [doi: 10.1007/978-3-319-98572-5_50]
63. Dobrovsky A, Borghoff UM, Hofmann M. Improving adaptive gameplay in serious games through interactive deep reinforcement learning. In: Klempous R, Nikodem J, Baranyi PZ, editors. Cognitive Infocommunications, Theory and Applications. Cham, Switzerland: Springer; 2019:411-432.
64. Sajjadi P, Ewais A, De Troyer O. Individualization in serious games: a systematic review of the literature on the aspects of the players to adapt to. Entertain Comput 2022 Mar;41:100468. [doi: 10.1016/j.entcom.2021.100468]

XSL•FO
RenderX

65. McCall R, Baillie L. Ethics, privacy and trust in serious games. In: Nakatsu R, Rauterberg M, Ciancarini P, editors. Handbook of Digital Games and Entertainment Technologies. Singapore, Singapore: Springer; 2020:1-30.
66. Caiani E. Ethics of digital health tools. European Society of Cardiology. 2020 Jul 8. URL: https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-18/ethics-of-digital-health-tools [accessed 2022-08-10]
67. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://apps.who.int/iris/bitstream/handle/10665/344249/9789240020924-eng.pdf [accessed 2022-08-09]

Abbreviations

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
SCT: social cognitive theory
WHO: World Health Organization

Edited by N Zary; submitted 27.04.22; peer-reviewed by J Gaspar, Z Reis, A Giovanelli; comments to author 19.05.22; revised version received 23.06.22; accepted 21.07.22; published 25.08.22.

Please cite as:
Martínez-Miranda J, Espinosa-Curiel IE
Serious Games Supporting the Prevention and Treatment of Alcohol and Drug Consumption in Youth: Scoping Review
JMIR Serious Games 2022;10(3):e39086
URL: https://games.jmir.org/2022/3/e39086
doi:10.2196/39086
PMID:36006694

©Juan Martínez-Miranda, Ismael Edrein Espinosa-Curiel. Originally published in JMIR Serious Games (https://games.jmir.org), 25.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Augmented Reality in Vascular and Endovascular Surgery: Scoping Review

Joshua Eves[1*], BSc, MBBCh BAO; Abhilash Sudarsanam[1], MSc, MBChB (Hons); Joseph Shalhoub[1,2], BSc, MBBS, MEd, FHEA, PhD; Dimitri Amiras[2,3*], BSc, MBBS

[1]Imperial Vascular Unit, Imperial College Healthcare NHS Trust, London, United Kingdom

[2]Department of Surgery & Cancer, Imperial College London, London, United Kingdom

[3]Department of Radiology, Imperial College Healthcare NHS Trust, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Joshua Eves, BSc, MBBCh BAO
Imperial Vascular Unit
Imperial College Healthcare NHS Trust
St Marys Hospital
Praed Street
London, W2 1NY
United Kingdom
Phone: 44 7881389569
Email: joshua.eves@nhs.net

Abstract

Background: Technological advances have transformed vascular intervention in recent decades. In particular, improvements in imaging and data processing have allowed for the development of increasingly complex endovascular and hybrid interventions. Augmented reality (AR) is a subject of growing interest in surgery, with the potential to improve clinicians' understanding of 3D anatomy and aid in the processing of real-time information. This study hopes to elucidate the potential impact of AR technology in the rapidly evolving fields of vascular and endovascular surgery.

Objective: The aim of this review is to summarize the fundamental concepts of AR technologies and conduct a scoping review of the impact of AR and mixed reality in vascular and endovascular surgery.

Methods: A systematic search of MEDLINE, Scopus, and Embase was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. All studies written in English from inception until January 8, 2021, were included in the search. Combinations of the following keywords were used in the systematic search string: ("augmented reality" OR "hololens" OR "image overlay" OR "daqri" OR "magic leap" OR "immersive reality" OR "extended reality" OR "mixed reality" OR "head mounted display") AND ("vascular surgery" OR "endovascular"). Studies were selected through a blinded process between 2 investigators (JE and AS) and assessed using data quality tools.

Results: AR technologies have had a number of applications in vascular and endovascular surgery. Most studies (22/32, 69%) used 3D imaging of computed tomography angiogram–derived images of vascular anatomy to augment clinicians' anatomical understanding during procedures. A wide range of AR technologies were used, with *heads up* fusion imaging and AR head-mounted displays being the most commonly applied clinically. AR applications included guiding open, robotic, and endovascular surgery while minimizing dissection, improving procedural times, and reducing radiation and contrast exposure.

Conclusions: AR has shown promising developments in the field of vascular and endovascular surgery, with potential benefits to surgeons and patients alike. These include reductions in patient risk and operating times as well as in contrast and radiation exposure for radiological interventions. Further technological advances are required to overcome current limitations, including processing capacity and vascular deformation by instrumentation.

(JMIR Serious Games 2022;10(3):e34501) doi:10.2196/34501

KEYWORDS

augmented reality; surgery; vascular; endovascular; head-mounted display; mobile phone

XSL•FO
RenderX

Introduction

Rationale

Emerging technologies are transforming vascular surgery at a rapid pace. In particular, the introduction of endovascular techniques has opened the way for a Cambrian explosion of technological evolution in terms of both hardware [1] and software [2].

One such technology is augmented reality (AR), which aims to enhance clinicians' ability by offering intuitive augmentation of the real environment with computer-generated real-time input of virtual information. AR lies on the continuum from virtual reality (VR), in which the user is immersed in a completely virtual setting, to real life. AR allows for minimal interaction between the virtual and real worlds, whereas mixed reality (MR) involves a combination of the real and virtual worlds where both elements are able to interact [3].

Complications in vascular surgery such as wound infections are associated with the extent of dissection, size of the wound, and duration of the surgery [4]. Anatomical localization with AR may help in reducing complications and improving overall outcomes. Although the "Getting It Right First Time" program used a principle of standardization across vascular surgery to improve outcomes in the United Kingdom [5], AR has promise in allowing for personalization taking into account individual patient anatomy and pathology while standardizing the technical approach by providing intraprocedural guidance.

Surgical AR uses a range of technologies, including everyday smartphone devices [6] and commercial products, including AR head-mounted displays (HMDs), which offer developers flexibility, allowing clinicians to experience virtual content that is overlaid directly onto the present reality [3].

Across a number of surgical specialties and subspecialties, the potential of AR as an important training tool has been identified [7]. With increasingly specialized surgical practice, more patients with comorbidities, and prevalent ethical challenges for surgical training, trainees have become increasingly reliant on simulation. Simulation has been demonstrated to effectively reduce training risks and costs [8]. Both VR and AR have been applied to surgical simulation successfully [9]; however, the superimposition of real-time information onto the real world and the flexibility offered by AR could make it a more realistic and adaptable simulation tool [10].

Literature on the applicability of AR in vascular surgery is extremely limited [11]. A recent review on the applicability of HMDs and smart glasses in vascular surgery highlighted the application of AR HMDs in a number of surgical specialties. However, only 4 papers in this review reported applications relevant to vascular surgery in particular [11]. To date, no review has been conducted on the application of the spectrum of AR technologies in the fields of vascular and endovascular surgery.

Objectives

This scoping review aimed to systematically search the literature, identify the current applications of AR in vascular and endovascular surgery, and summarize key results and learning points while identifying avenues and gaps for future research in this evolving field.

Methods

Protocol

A scoping review was chosen after advice from PROSPERO (international prospective register of systematic reviews), and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [12] was used to ensure validity.

Eligibility Criteria

The inclusion criteria were (1) studies in the English language, (2) a minimum of level V evidence using the Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence, (3) use of AR in vascular or endovascular surgery, and (4) applicability to clinical practice or training reported.

The exclusion criteria were (1) review articles or conference abstracts, (2) non–English-language articles, (3) articles lacking an available full text, (4) use of AR outside of vascular or endovascular surgery, and (5) use only of VR.

Search and Information Sources

A systematic search of the MEDLINE, Scopus, and Embase databases was performed, allowing for access to a range of global clinical, scientific, and engineering research. All studies written in English from inception until January 8, 2021, were included in the search.

The following keywords were used in the systematic search string for all 3 database searches: ("augmented reality" OR "hololens" OR "image overlay" OR "daqri" OR "magic leap" OR "immersive reality" OR "extended reality" OR "mixed reality" OR "head mounted display") AND ("vascular surgery" OR "endovascular"). No limits were applied.

Selection of Sources of Evidence

First, a blinded and independent process of selection based on titles and abstracts was performed without collusion by 2 authors (AS and JE), with a third author (JS) consulted with regard to discrepancies. Next, a selection of eligible studies was conducted by analyzing the full texts.

Data Items

Each study was evaluated individually, and the following variables were sought for data collection: year of study, category (endovascular surgery [aortic, peripheral, venous, or visceral], open surgery, and training), applied study or concept and design, number of patients (if the study was clinically applied), risk of bias (see the following section for the tools used), characteristics of the study group and operators, methods and outcomes, imaging type, and type of display (if applicable).

Critical Appraisal of Individual Sources of Evidence

The quality of the data was evaluated using the Cochrane tools Risk of Bias 2 for randomized trials and Risk of Bias In Non-randomized Studies of Interventions for nonrandomized trials [13,14]. For those studies that included human participants but did not meet the criteria for the aforementioned scoring

XSL•FO
RenderX

tools, the Johanna Briggs Institute Critical Appraisal Checklist was used, with a score derived based on the checklist. A score of 1 was assigned for *Yes*, a score of 0.5 was assigned for *unclear*, and a score of 0 was assigned for *no*. The results were presented as percentages. A higher percentage represented a reduced risk of bias.

Data Charting Process and Synthesis

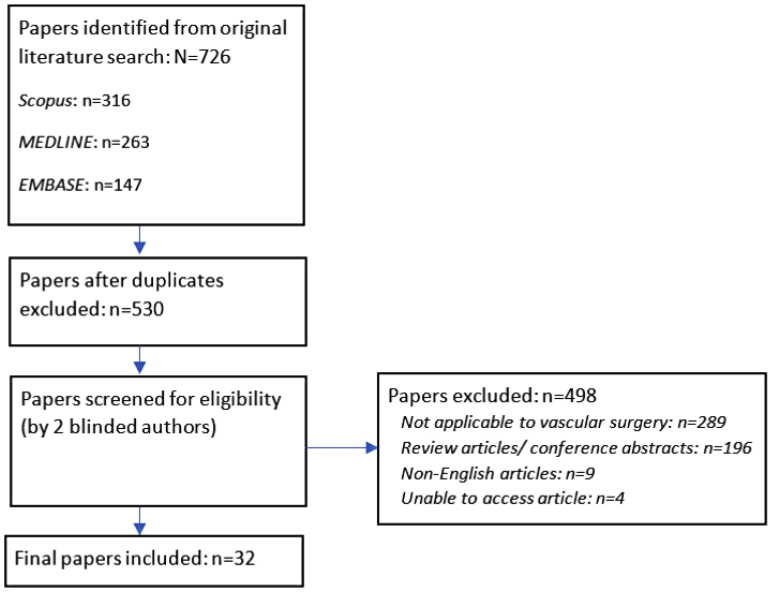
Data were extracted from eligible studies into evidence tables to summarize the following: year of publication, type of study design, number of patients, method, and outcome. A second table presents more technical aspects of the research, including the type of imaging, methods for tracking or registration, and display. The data collected in the evidence tables were used to define the main themes of discussion. Any data related to the application of AR in vascular and endovascular surgery could be synthesized.

Results

Selection of Sources of Evidence

A total of 726 articles were identified from the initial search across the 3 databases, with 32 (4.4%) meeting the inclusion and exclusion criteria and being included in the final results. Please see Figure 1 for detailed information, including the reasons for article exclusion.

Figure 1. Search strategy. Search of MEDLINE, Scopus, and Embase databases. Exclusions shown.



Characteristics of the Sources of Evidence

A total of 32 articles applicable to vascular and endovascular surgery met the criteria for inclusion. The included articles were sorted into three categories: open vascular surgery, endovascular surgery (subdivided into aortic, visceral, peripheral, and venous, where applicable), and training (Multimedia Appendix 1 [6,15-45]). Many studies were preclinical in application, with 44% (14/32) including human study populations with limited numbers. There were no human studies with equivalent outcome measures for comparative quantitative data analysis.

Critical Appraisal

Quality assessment could only be performed in cases where human study populations were evaluated, which was in 44% (14/32) of the studies. The Cochrane risk of bias assessment was not used as no randomized trials were included in our review. Among the clinical studies, 6% (2/32) were cohort studies [15,19], which were found to have a moderate risk of bias according to the Cochrane risk of bias tool for nonrandomized trials [13]. Of the 32 studies, 1 (3%) was a case-control study [25] with a score of 55% according to the Joanna Briggs Institute Critical Appraisal Checklist, 7 (22%) were case series studies [28,32,37-40] with SD (16.1%) an average score of 70% (SD 16.1%; range 50%-80%), and 3 (9%) were case reports [16,36,39] an average score of 65% (range 40%-95%).

The articles were further subdivided into studies that focused on *concept and design* to reflect the early stage of this technology in vascular surgery and those that conducted research directly *applied* to surgery, training, or simulation.

Synthesis of Results

Overview

The proportion of studies related to concept development of AR was 38% (12/32), whereas the rest were applied to clinical and simulation contexts. Of the 32 studies, 29 (91%) involved deriving digital structures of blood vessels from clinical imaging (Multimedia Appendix 2 [6,15-45]), including noncontrast computed tomography (CT), CT angiography, cone-beam CT (CBCT), magnetic resonance angiography, conventional ultrasound (US), intravascular US, and 3D US.

Virtual data are processed into a virtual object and displayed to the clinician on a screen, projector, monitor, or AR HMD. Several steps are required to complete processing, including segmentation to reduce the model to the area of interest (eg, aorta), often by thresholding via a Hounsfield scale, for example, and then using a 3D computer-generated surface mesh, refining the mesh by filling holes or defects into a model and sometimes

XSL•FO
RenderX

using a slicer to break up the model into its different anatomical sections [41].

Registration and Tracking

Registration and tracking are essential components of AR, allowing for the alignment of virtual and real data in a usable way. Registration is commonly performed using markers, where a particular pattern or anatomical landmark in the real world is used as a reference corresponding to a virtual marker derived from medical imaging [10]. This can be performed manually or using trackers. Marker-less tracking is now possible owing to technological advances that correlate patterns in real and virtual data in real time [46,47]. Integrated or external optical sensors (infrared or color) can be used to track markers or recognize patterns from a patient's anatomy. In recent years, Red, Green, Blue-Depth cameras that track both depth and color simultaneously have allowed for the contemporaneous tracking of the real world [48]. Electromagnetic tracking is a useful method for tracking instruments deep below the surface [26]. In many cases, such as with Microsoft HoloLens, a variety of sensor inputs are used for hybrid tracking [49].

Challenges with these techniques include the deformability of the anatomy; for example, when stiff endovascular devices are inserted into blood vessels. These changes can cause registration errors, which can be clinically significant, for example, with complex visceral anatomy in endovascular aneurysm repair (EVAR). Groher et al [24] proposed computational algorithms to allow for some accurate deformability of 3D models for registration with 2D fluoroscopy. Another innovative tracking solution combating this problem involves electromagnetic trackers on the tips of angiography catheters, which can provide accurate information on the 3D position and orientation of the catheter in space, such as those used by Garcia-Vasquez et al [27].

Alternatively, intravascular US on the tip of catheters can be used to create a 3D model of blood vessels during endovascular procedures with magnetic trackers to orientate the catheter in real time, as proposed by Shi et al [29]. Parrini et al [21] tested the use of freehand external magnets as a means of guiding magnetic endovascular devices to their targets using AR vessel models on a phantom.

Endovascular Procedures

Given the inherent reliance of endovascular procedures on imaging, the potential benefits of AR were identified early, in particular to improve clinicians' understanding of 3D anatomy during critical steps as well as to reduce time and radiation and contrast exposure [50]. Image overlay techniques are used for complex endovascular work worldwide and usually rely on merging live fluoroscopy with 2D or 3D images from x-ray, CT, or magnetic resonance imaging.

EVAR and Aortic Disease

EVAR is increasingly being performed for the repair of both complex and conventional infrarenal abdominal aortic aneurysms in the elective and emergency settings.

AR visualization has now been used in planning and EVAR navigation, with a case report by Rynio et al [16] describing the use of an AR HMD to project an aortic aneurysm and bones (vertebral column lying posterior to the aneurysm) as a 3D hologram that responds to gestures and voice commands.

A "3D road map" published in 2013 was developed by Fukuda et al [34] using preoperative CT imaging, where bone marrow subtraction and the use of the iliac crest and lumbar vertebrae as landmarks resulted in the creation of an image overlay to guide aortic endografting. This reduced the need for digital subtraction angiography (DSA) and its associated risks, ultimately giving rise to fusion imaging.

Fusion imaging has been used for endovascular navigation in hybrid vascular operating theaters, particularly in complex fenestrated EVARs. This uses preoperative 3D CT imaging to overlay vascular structures onto perioperative 2D fluoroscopy images with the aim of reducing fluoroscopy time, contrast agent dosing, and overall operating time. This can also be used in a conventional operating theater with a mobile C-arm with good results for conventional aortic endografting, as reported by Kaladji et al [25]. Koutouzi et al [32] modified this modality to require carbon dioxide DSA to confirm accurate registration of the CT-derived 3D vascular overlay to avoid iodinated contrast in patients with poor renal function or contrast allergies.

Kaladji et al [33] went on to successfully demonstrate the use of fusion imaging to perform EVAR without the use of any iodinated contrast in the pre- or perioperative phase in a case series of 6 patients. Patients with severe chronic kidney disease underwent unenhanced CT imaging. Centerlines for EVAR planning were manually drawn, and key anatomical points or landing zones were marked following segmentation and processing. A preoperative 3D image overlay reconstruction with these markers was then projected onto 2D fluoroscopy imaging to guide the placement of the aortic endograft without the use of a contrast agent (Figure 2). No endoleaks were noted during postoperative duplex surveillance, with minimal error in the positioning of the aortic endograft on postoperative CT imaging and no significant reduction in renal function.

A series of 101 patients by Schulz et al [19] in 2016 tested the use of image overlay alone in patients undergoing conventional infrarenal EVAR. Fusion image overlay using images from preoperative CT angiograms with images obtained from an intraoperative CBCT was compared with intraoperative DSA. Although <5-mm accuracy was observed in most patients (68% of the patient cohort), significant deviations were noted in some patients, including significant caudal deviation in 9 patients (9% of the patient cohort), which would have resulted in coverage of the lowest-lying renal artery, resulting in significant operative morbidity. Therefore, a contrast injection study for cannulation of the lowest-lying visceral vessel was recommended.

Koutouzi et al [28] described a technique that aims to use image overlay to reduce the risk of covering intercostal arteries during thoracic EVAR (TEVAR), which may affect spinal cord perfusion, causing paraparesis. Intraoperative CBCT was used to perform 3D-3D registration with a preoperative CT angiogram to enhance accuracy. Once calibrated onto live fluoroscopy, this overlaid 3D model was used to guide the TEVAR and avoid

XSL•FO
RenderX

preplanned intercostal arteries. A case series of 7 was reported without spinal cord injury.

A common problem with image overlay techniques is the nondeformability of a rigid 3D model created from a preoperative CT scan. Garcia-Vasquez et al [27] attempted to overcome this problem using 3D US to create a real-time model of the patients' anatomy that was visible to a clinician wearing the AR HMD. Coupled with the electromagnetic catheter tip tracking system used by von Haxthausen et al [26], this concept has been proposed to perform EVAR completely free of radiation and intravenous contrast and tested on a phantom model with a 3D model of an aortic aneurysm [27]. The concept showed promising progress (Figure 3); however, issues around registration accuracy remain.

Figure 2. During the procedure, different information was overlaid onto the native 2D fluoroscopic image to guide instrument placement: (A) none, (B) centerlines and key points can be projected, and (C) artificially enhanced aortic volume. Reprinted from European Journal of Vascular & Endovascular Surgery, 49/3, Kaladji A, Dumenil A, Mahé G, Castro M, Cardon A, Lucas A, Haigron P, Safety and accuracy of endovascular aneurysm repair without pre-operative and intra-operative contrast agent, 255-261, 2015, with permission from Elsevier [33].

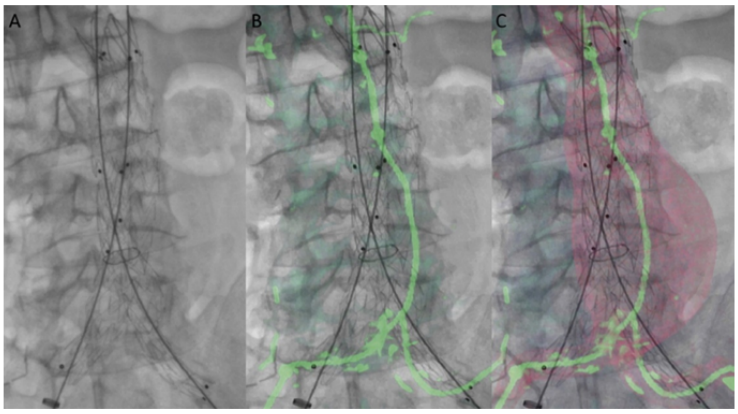


Figure 3. View from HoloLens display of 3D vasculature from 3D ultrasound being projected onto a phantom. Reproduced from García-Vázquez V et al [27] under the Creative Commons Attribution Non-Commercial License.



Endovascular—Peripheral Arterial Disease

Endovascular techniques are being applied to increasingly complex surgical problems in patients with comorbidities. The potential to reduce radiation and contrast exposure while navigating tortuous vessels makes the application of AR technology more attractive.

Lu et al [36] described successful retrograde peroneal access through an AR system with CT angiogram image–derived overlay of fluoroscopic images. AR glasses were worn by the operator to guide the needle trajectory. The case report highlighted technical difficulties, including the need for manual registration and having to align virtual dots on their AR HMD with fiducial markers on the patient's leg. In addition, aligning the trajectory of the needle path on the AR HMD required the acquisition of new technical skills.

The work of Goudeketting et al [37] focused on the accuracy of image fusion based on preprocedural contrast-enhanced magnetic resonance angiography during percutaneous angioplasty of iliac lesions. They found that guidewires and endovascular catheters did not cause significant vessel displacement to influence image fusion according to angiographic experts.

Swerdlow et al [15] retrospectively compared carotid stenting with 2D-3D image fusion (46 patients) and without (70 patients). Magnetic resonance imaging or CT angiography images were overlaid onto the real-time 2D fluoroscopy. They observed significantly improved cannulation times and reduced radiation exposure using this technique.

Endovascular—Venous Disease

Endovascular venous intervention, compared with arterial intervention, can be more challenging because of the lack of

XSL•FO
RenderX

landmarks from vessel calcification and reduced vessel wall thickness. Schwein et al [38] used image fusion techniques with magnetic resonance venography to successfully recanalize 4 patients with central venous occlusion. It was felt that magnetic resonance venography image fusion improved clinician confidence and the safety of difficult venous endovascular navigation.

Open Surgery

Overview

"Traditional" open vascular surgery has increasingly been replaced by novel endovascular techniques, which, in combination with problems with tissue deformation, as previously described, may explain why AR is less well researched in this field.

The use of a new image overlay system has been demonstrated successfully in lower limb revascularization surgery by Mochazuki et al [40]. This system used preoperative CT imaging to help locate the target distal anastomotic site and plan a limited incision accordingly using a branch artery as a reference point (Figure 4).

Modern-day mobile phones have multiple sensors that may be used for AR-assisted surgery. The benefits of easy accessibility (most operators have access to a mobile phone) and cost make them an attractive option. An AR-assisted surgery system was developed by Aly [6] to provide 3D guidance during vascular procedures. CT angiogram–derived images were used by a smartphone to produce a 3D guidance model fused with the patient's anatomy, and rotational and positional tracking were compared for planning an iliofemoral bypass as well as for an endovascular sheath placement in the common femoral artery (Figure 5).

Figure 4. Computed tomography–derived 3D model is superimposed onto camera images. Doughnut-shaped fiducial markers on the limb help match images so that the target artery and its branches can be overlaid onto the limb as red (A) or blue (B) lines. Reprinted by permission from the Springer Nature Customer Service Centre GmbH: Springer Nature. New simple image overlay system using a tablet PC for pinpoint identification of the appropriate site for anastomosis in peripheral arterial reconstruction. Mochizuki Y, Hosaka A, Kamiuchi H, Nie JX, Masamune K, Hoshina K, et al. 2016 [40].

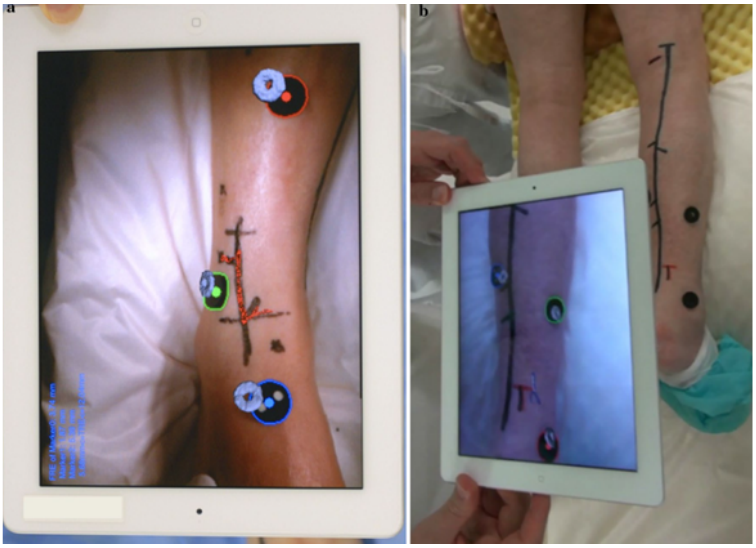
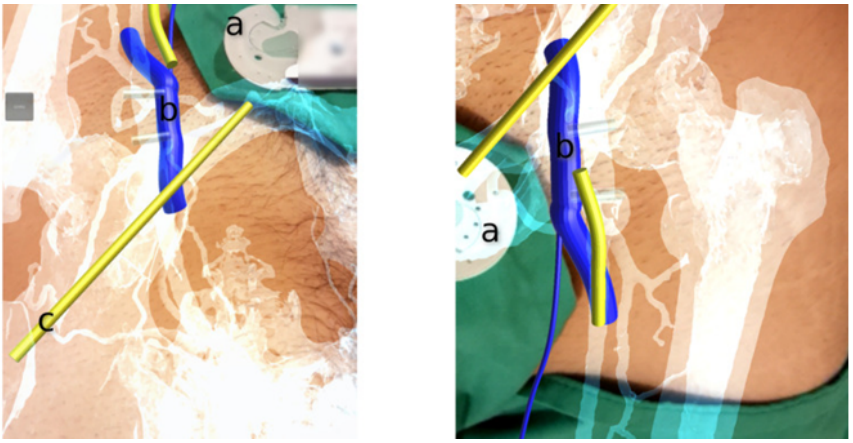


Figure 5. Screen capture of augmented reality–assisted surgery system developed by Aly [6]. Fiducial marker (sterile suture pack) used for registration and tracking with 3D model derived from computed tomography angiogram superimposed from mobile phone. (A) Tracking marker, (B) common femoral vein, and (C) inguinal ligament. Reproduced under the Creative Commons Attribution License CC-BY 4.0.



XSL•FO
RenderX

Robotic Surgery

The use of AR with robot-assisted surgery is of increasing academic interest worldwide. Pietrabissa et al [39] successfully used an AR HMD to superimpose a surgical anatomy from CT angiography onto the patient's body for preoperative planning to guide trocar placement and dissection so as to minimize disruption using the da Vinci surgical system. Although robot-assisted surgery is becoming increasingly common for vascular interventions, the use of AR is in the early stages of adoption and investigation.

Training and Simulation

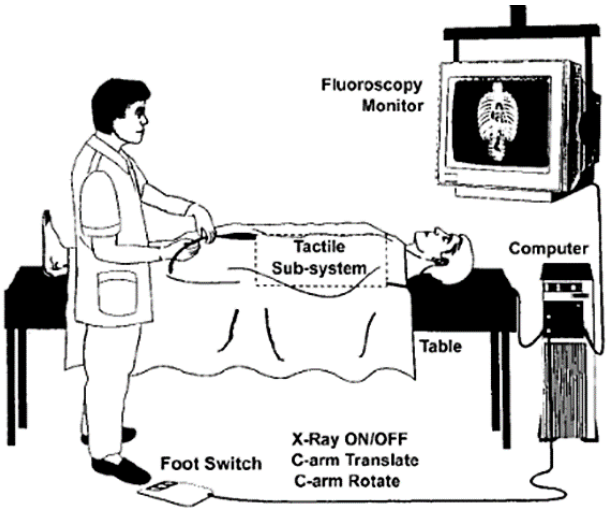
Published studies on AR vascular and endovascular training were limited in our review to 16% (5/32) of the papers. Of these 5 papers, 1 (20%) applied this to training vascular surgeons, whereas 4 (80%) focused on design concepts.

Mangina et al [41] tested the concept of creating a 3D model of the aorta from CT angiography images using a number of programming tools and platforms, including *Unity*, *ARToolKit*, and *Pro/ENGINEER*. A model aorta was created with the potential for merging with VR and AR tracking technologies to create an accurate training and educational tool for application to both open and endovascular training.

Bartesaghi et al [43] used similar techniques to create a 3D model of the aorta projected onto a mannequin to allow for 3D AR simulation of an EVAR. The original model was derived from CT angiography images, and segmentation with meshing was performed to create the model. The software was optimized for the interplay between the simulation and a hardware mannequin, allowing surgical trainees to learn the steps of an EVAR.

Similarly, Anderson et al [45] described the development of a PC-based AR simulation system for interventional radiology (Figure 6). This system allows for real-time manipulation of catheters and guidewires through a tactile user interface device that creates hand-eye coordinated realism of the movement of instruments over 3D-generated models of vessels. Extra enhancements have been built in, including the administration and subsequent washout of contrast and control of the C-arm. The paper reports good initial feedback from neurological and peripheral vascular interventionalists who felt it had 75% of the desired functionality and was useful for educational simulation, with further work required to achieve sufficient realism for preoperative planning.

Figure 6. Early augmented reality interventional simulator design. Reproduced from Anderson et al [45], with permission from © Georg Thieme Verlag KG. Note: The permission of the figure stays with the publisher, and any further reuse will need explicit permission from the publisher.



Crowley et al [42,51] built on the work of Mangina et al [41], further exploring the concept of EVAR MR simulation. They found that the creation of a solid 3D-printed model of an aortic aneurysm through which real instruments (guidewires, catheters, and delivery devices) can be placed and manipulated gave the simulation tactile realism. A mobile device provides the user with fluoroscopic data through an AR software (ARToolKit or Daqri) while haptic feedback is provided by the 3D model. Limitations of 3D-printed models include the lack of vascular wall deformability and the realism of feedback from catheter tips against the model wall. The system's usability was tested by medical students and university lecturers.

An AR system for US-guided endovascular surgical training was devised by Cheng et al [20]. Live US images were transferred to a prescanned phantom model of the aorta and iliac arteries, which was extended to a computer-assisted remote endovascular surgery model. The user is able to navigate an endovascular catheter using a robotic device with computer-generated images to augment the 3D understanding of the catheter in space. The phantom provides both haptic and visual feedback aiming to be as realistic as possible.

Rudarakanchana et al [44] applied the concept of simulation to the pressured setting of ruptured EVAR. This was achieved through VR simulators being integrated into a simulated angio suite with the whole team present. A total of 10 teams were tested: 5 led by interventional specialists and 5 led by trainees. This study measured the time to achieve proximal control and total procedure and fluoroscopy times. Experts were significantly faster than trainees using reduced fluoroscopy time, suggesting that the simulation model had good applicability to real-life experience. In particular, the value of the simulation for improving team communication and human

XSL•FO
RenderX

factor skills was highlighted through feedback, suggesting that AR and MR simulation could be useful adjuncts to improving ruptured EVAR outcomes.

Our review demonstrates that the most used modality for obtaining data for display is CT angiography (Table 1), with CBCT often used in the registration process with the patient in the operating suite. The type of display will depend on the application, and a range of display types were included, with image fusion studies making monitors the most prevalent display (Table 2).

Table 1. Most of the included studies derived data from 3D clinical imaging, which can be overlaid onto the real world (N=32).

Data source	Studies, n (%)
Noncontrast CT[a]	1 (3)
CT angiography	22 (69)
Cone-beam CT	11 (34)
MR[b] angiography or venography	3 (9)
Conventional US[c]	1 (3)
Intravascular US	1 (3)
3D US	2 (6)
Rotational XR[d]	1 (3)
Computer-generated model	3 (9)

[a]CT: computed tomography.

[b]MR: magnetic resonance.

[c]US: ultrasound.

[d]XR: x-ray.

Table 2. Types of augmented reality (AR) displays. A monitor was the most frequently used type of display, often used for image fusion in interventional procedures (N=32).

Types of AR display	Studies, n (%)
AR HMD[a] (including HoloLens)	9 (28)
Monitor	20 (63)
Mobile phone	2 (6)
Projector	1 (3)

[a]HMD: head-mounted display.

Discussion

Summary of Evidence

The reviewed studies demonstrate that AR in surgery often relies on the registration of a virtual image or object onto the real patient using a tracking method, which can trace the real environment and place virtual objects in the correct position and orientation. Our study revealed a wide range of applications of AR in vascular surgery. Image overlay technology is increasingly used, with its value being reported in endovascular treatment of peripheral arterial disease [37], deep venous disease [38], carotid stenting [15], and aortic disease (EVAR [19,32], TEVAR [28], and complex EVAR cases [31,52]). There is evidence that it can improve technical success while reducing procedural times, radiation dose, and contrast volume.

However, concerns remain regarding the accuracy of image registration, especially with more complex anatomies [19]. A common problem is the deformation of vascular structures with respiration, surgical manipulation, or *stretching out* of vessels (such as iliac and target visceral arteries for EVAR) with rigid stent delivery systems. Emerging technologies offer novel solutions for improving the real-time accuracy of overlaid data. Robust deformable registration algorithms or intraoperative real-time 3D scanners are potential solutions to assist AR-guided surgery on soft or deformable structures. In particular, 3D and intravascular US offer radiation- and contrast-free modalities to create real-time images for intraoperative use [27], whereas electromagnetic tracking provides contemporaneous spatial information about catheter devices [26]. The integration of these technologies demands high levels of data transmission and processing, which must be met by technology to prevent lag [53].

AR HMDs allow the wearer hands-free access to virtual data overlaid directly onto the patient. Several studies with EVAR in particular have shown early promise that they can reduce the need for radiation and contrast exposure while improving accuracy [26,27,31]. Studies on open and robotic surgery corroborate this [39,54]. AR HMDs such as HoloLens still have a few technical concerns that could restrict their usefulness.

XSL•FO
RenderX

Their weight (500-645 g depending on the device and manufacturer) can cause discomfort and fatigue during long procedures. AR HMD processing capacity and memory are limited, which can restrict certain applications. The immersive experience can be limited by a restricted field of view and projection size, whereas image quality and computational time will improve as the technology evolves. There is no consensus yet on the optimal methods for merging and displaying virtual and real information so that depth perception or focus and visual clutter do not distract the wearer [11,55].

Most of the included studies (28/32, 88%) described small-scale implementations of AR with limited study participants and a wide range of AR hardware and software platforms in their research. Only 44% (14/32) of the studies included human study populations. As AR becomes more ubiquitous for vascular surgery, more evidence will become available, and we believe that these limitations will pose less of an issue.

A significant proportion of the studies were preclinical in their applications and demonstrated proof-of-concept findings. Nevertheless, AR solutions are applicable to a wide spectrum of fields in vascular surgery, including those highlighted in this scoping review.

Limitations

Three databases (MEDLINE, Embase, and Scopus) were searched for this review, allowing for a broad assessment of the current impact of AR clinically and educationally while also scoping potential future applications. However, in this fast-moving field, AR applications under development by private enterprises were not included in the contents of our review. Although our selection criteria focused on vascular and endovascular surgery, there is significant crossover with work applied to other medical and surgical specialties, and wider development within interventional radiology and allied surgical specialties will further inform applications within vascular and endovascular surgery.

Conclusions

AR has shown potential to enhance accuracy and reduce procedural time, radiation exposure, and contrast dose in a range of vascular surgery applications by digitally augmenting the clinicians' procedural ability, reducing surgical risk, and improving patient outcomes. Clinicians, who demand high levels of accuracy and patient safety, should be aware of potential technical pitfalls when using AR technology on patients but should also be aware of the potential benefits. AR has also been shown to be an increasingly valuable tool in surgical simulation and education. As technology improves, it can be expected that AR will become increasingly relied upon as an arrow in the surgical quiver and that more applications will be found for AR, which could benefit clinicians. Future development and studies should assess whether the use of AR affords improvements in patient experience and in clinical effectiveness as objective measures of improvement in outcomes and cost-effectiveness of this technology.

Conflicts of Interest

JS received grant funding from the UK National Institute for Health Research and the British Heart Foundation, as well as consulting with Oxford HealthTech Ltd. All the funding was paid to the institution and is not related to this publication. DA is a clinical advisor for Medical iSight Ltd.

Multimedia Appendix 1
Studies applicable to augmented reality in vascular and endovascular surgery.
[DOCX File , 28 KB - games_v10i3e34501_app1.docx]

Multimedia Appendix 2
Imaging, registration, and display types within review.
[DOCX File , 21 KB - games_v10i3e34501_app2.docx]

References

1. Cheng S. Novel endovascular procedures and new developments in aortic surgery. Br J Anaesth 2016 Sep;117 Suppl 2:ii3-i12 [FREE Full text] [doi: 10.1093/bja/aew222] [Medline: 27566806]
2. Bedi VS, Sharma N. Game changers in vascular and endovascular surgery. Indian J Vasc Endovasc Surg 2020;7:13-17. [doi: 10.4103/IJVES.IJVES_25_20]
3. Rahman R, Wood ME, Qian L, Price CL, Johnson AA, Osgood GM. Head-mounted display use in surgery: a systematic review. Surg Innov 2020 Mar 12;27(1):88-100. [doi: 10.1177/1553350619871787] [Medline: 31514682]
4. Gwilym BL, Saratzis A, Benson R, Forsythe R, Dovell G, Dattani N, et al. Groin wound infection after vascular exposure (GIVE) multicentre cohort study. Int Wound J 2020 Nov 25;18(2):164-175. [doi: 10.1111/IWJ.13508]
5. Gray WK, Day J, Horrocks M. Editor's choice - volume-outcome relationships in elective abdominal aortic aneurysm surgery: analysis of the uk hospital episodes statistics database for the getting it right first time (GIRFT) programme. Eur J Vasc Endovasc Surg 2020 Oct;60(4):509-517 [FREE Full text] [doi: 10.1016/j.ejvs.2020.07.069] [Medline: 32807679]
6. Aly O. Assisting vascular surgery with smartphone augmented reality. Cureus 2020 May 08;12(5):e8020 [FREE Full text] [doi: 10.7759/cureus.8020] [Medline: 32528759]

XSL•FO
RenderX

7. Loukas C. Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators. Int J Advanced Robot Auto 2016 Jul 11;1(2):1-9. [doi: 10.15226/2473-3032/1/2/00109]
8. Aggarwal R, Mytton OT, Derbrew M, Hananel D, Heydenburg M, Issenberg B, et al. Training and simulation for patient safety. Qual Saf Health Care 2010 Aug 06;19 Suppl 2(Suppl 2):i34-i43 [FREE Full text] [doi: 10.1136/qshc.2009.038562] [Medline: 20693215]
9. McKnight RR, Pean CA, Buck JS, Hwang JS, Hsu JR, Pierrie SN. Virtual reality and augmented reality-translating surgical training into surgical technique. Curr Rev Musculoskelet Med 2020 Dec 11;13(6):663-674 [FREE Full text] [doi: 10.1007/s12178-020-09667-3] [Medline: 32779019]
10. Lungu AJ, Swinkels W, Claesen L, Tu P, Egger J, Chen X. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. Expert Rev Med Devices 2021 Jan 16;18(1):47-62. [doi: 10.1080/17434440.2021.1860750] [Medline: 33283563]
11. Lareyre F, Chaudhuri A, Adam C, Carrier M, Mialhe C, Raffort J. Applications of head-mounted displays and smart glasses in vascular surgery. Ann Vasc Surg 2021 Aug;75:497-512. [doi: 10.1016/j.avsg.2021.02.033] [Medline: 33823254]
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
13. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
14. Mayhew AD, Kabir M, Ansari MT. Considerations from the risk of bias perspective for updating Cochrane reviews. Syst Rev 2015 Oct 06;4(1):136 [FREE Full text] [doi: 10.1186/s13643-015-0122-3] [Medline: 26445323]
15. Swerdlow NJ, Jones DW, Pothof AB, O'Donnell TF, Liang P, Li C, et al. Three-dimensional image fusion is associated with lower radiation exposure and shorter time to carotid cannulation during carotid artery stenting. J Vasc Surg 2019 Apr;69(4):1111-1120 [FREE Full text] [doi: 10.1016/j.jvs.2018.07.038] [Medline: 30301693]
16. Rynio P, Witowski J, Kamiński J, Serafin J, Kazimierczak A, Gutowski P. Holographically-guided endovascular aneurysm repair. J Endovasc Ther 2019 Aug 13;26(4):544-547. [doi: 10.1177/1526602819854468] [Medline: 31190631]
17. Grinshpoon A, Sadri S, Loeb GJ, Elvezio C, Feiner SK. Hands-free interaction for augmented reality in vascular interventions. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018 Presented at: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); Mar 18-22, 2018; Tuebingen/Reutlingen, Germany.
18. Kilian-Meneghin J, Xiong Z, Guo C, Rudin S, Bednarek DR. Evaluation of methods of displaying the real-time scattered radiation distribution during fluoroscopically-guided interventions for staff dose reduction. Proc SPIE Int Soc Opt Eng 2018 Feb;10573:1057366 [FREE Full text] [doi: 10.1117/12.2294575] [Medline: 30026638]
19. Schulz CJ, Schmitt M, Böckler D, Geisbüsch P. Fusion imaging to support endovascular aneurysm repair using 3D-3D registration. J Endovasc Ther 2016 Oct 28;23(5):791-799. [doi: 10.1177/1526602816660327] [Medline: 27456083]
20. Cheng I, Shen R, Moreau R, Brizzi V, Rossol N, Basu A. An augmented reality framework for optimization of computer assisted navigation in endovascular surgery. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 26-30, 2014; Chicago, IL, USA. [doi: 10.1109/embc.2014.6944908]
21. Parrini S, Cutolo F, Freschi C, Ferrari M, Ferrari V. Augmented reality system for freehand guide of magnetic endovascular devices. In: Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 26-30, 2014; Chicago, IL, USA. [doi: 10.1109/embc.2014.6943635]
22. van den Berg JC. Update on new tools for three-dimensional navigation in endovascular procedures. Aorta (Stamford) 2014 Dec;2(6):279-285 [FREE Full text] [doi: 10.12945/j.aorta.2014.14-054] [Medline: 26798746]
23. Wang J, Fallavollita P, Wang L, Kreiser M, Navab N. Augmented reality during angiography: integration of a virtual mirror for improved 2D/3D visualization. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2012 Presented at: IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Nov 05-08, 2012; Atlanta, GA, USA. [doi: 10.1109/ISMAR.2012.6402565]
24. Groher M, Zikic D, Navab N. Deformable 2D-3D registration of vascular structures in a one view scenario. IEEE Trans Med Imaging 2009 Jun;28(6):847-860. [doi: 10.1109/tmi.2008.2011519]
25. Kaladji A, Villena A, Pascot R, Lalys F, Daoudal A, Clochard E, et al. Fusion imaging for EVAR with mobile C-arm. Ann Vasc Surg 2019 Mar;55:166-174. [doi: 10.1016/j.avsg.2018.06.006] [Medline: 30092429]
26. von Haxthausen F, Jäckle S, Strehlow J, Ernst F, García-Vázquez V. Catheter pose-dependent virtual angioscopy images visualized on augmented reality glasses. Curr Direction Biomed Eng 2019;5(1):289-291. [doi: 10.1515/cdbme-2019-0073]
27. García-Vázquez V, von Haxthausen F, Jäckle S, Schumann C, Kuhlemann I, Bouchagiar J, et al. Navigation and visualisation with HoloLens in endovascular aortic repair. Innov Surg Sci 2018 Sep;3(3):167-177 [FREE Full text] [doi: 10.1515/iss-2018-2001] [Medline: 31579781]
28. Koutouzi G, Sandström C, Skoog P, Roos H, Falkenberg M. 3D image fusion to localise intercostal arteries during TEVAR. EJVES Short Rep 2017;35:7-10 [FREE Full text] [doi: 10.1016/j.ejvssr.2017.03.001] [Medline: 28856332]

XSL•FO
RenderX

29. Shi C, Tercero C, Ikeda S, Ooe K, Fukuda T, Komori K, et al. In vitro three-dimensional aortic vasculature modeling based on sensor fusion between intravascular ultrasound and magnetic tracker. Int J Med Robot 2012 Sep 01;8(3):291-299. [doi: 10.1002/rcs.1416] [Medline: 22298412]
30. Rolls AE, Maurel B, Davis M, Constantinou J, Hamilton G, Mastracci TM. A comparison of accuracy of image- versus hardware-based tracking technologies in 3D fusion in aortic endografting. Eur J Vasc Endovasc Surg 2016 Sep;52(3):323-331 [FREE Full text] [doi: 10.1016/j.ejvs.2016.05.001] [Medline: 27389943]
31. Duménil A, Kaladji A, Castro M, Göksu C, Lucas A, Haigron P. A versatile intensity-based 3D/2D rigid registration compatible with mobile C-arm for endovascular treatment of abdominal aortic aneurysm. Int J Comput Assist Radiol Surg 2016 Sep 26;11(9):1713-1729. [doi: 10.1007/s11548-016-1416-1] [Medline: 27230779]
32. Koutouzi G, Henrikson O, Roos H, Zachrisson K, Falkenberg M. EVAR guided by 3D image fusion and CO2 DSA: a new imaging combination for patients with renal insufficiency. J Endovasc Ther 2015 Dec 17;22(6):912-917. [doi: 10.1177/1526602815605468] [Medline: 26384396]
33. Kaladji A, Dumenil A, Mahé G, Castro M, Cardon A, Lucas A, et al. Safety and accuracy of endovascular aneurysm repair without pre-operative and intra-operative contrast agent. Eur J Vasc Endovasc Surg 2015 Mar;49(3):255-261 [FREE Full text] [doi: 10.1016/j.ejvs.2014.12.003] [Medline: 25617257]
34. Fukuda T, Matsuda H, Doi S, Sugiyama M, Morita Y, Yamada M, et al. Evaluation of automated 2D-3D image overlay system utilizing subtraction of bone marrow image for EVAR: feasibility study. Eur J Vasc Endovasc Surg 2013 Jul;46(1):75-81 [FREE Full text] [doi: 10.1016/j.ejvs.2013.04.011] [Medline: 23642524]
35. Carrell TW, Modarai B, Brown JR, Penney GP. Feasibility and limitations of an automated 2D-3D rigid image registration system for complex endovascular aortic procedures. J Endovasc Ther 2010 Aug;17(4):527-533. [doi: 10.1583/09-2987MR.1] [Medline: 20681771]
36. Lu W, Wang L, Zhou W, Fu W. Augmented reality navigation to assist retrograde peroneal access for the endovascular treatment of critical limb ischemia. J Vasc Surg Cases Innov Tech 2019 Dec;5(4):518-520 [FREE Full text] [doi: 10.1016/j.jvscit.2019.06.004] [Medline: 31768482]
37. Goudeketting SR, Heinen SG, van den Heuvel DA, van Strijen MJ, de Haan MW, Slump CH, et al. The use of 3D image fusion for percutaneous transluminal angioplasty and stenting of iliac artery obstructions: validation of the technique and systematic review of literature. J Cardiovasc Surg 2018 Feb;59(1):26-36. [doi: 10.23736/s0021-9509.17.10224-7]
38. Schwein A, Lu T, Chinnadurai P, Kitkungvan D, Shah D, Chakfe N, et al. Magnetic resonance venography and three-dimensional image fusion guidance provide a novel paradigm for endovascular recanalization of chronic central venous occlusion. J Vasc Surg Venous Lymphat Disord 2017 Jan;5(1):60-69. [doi: 10.1016/j.jvsv.2016.07.010] [Medline: 27987612]
39. Pietrabissa A, Morelli L, Ferrari M, Peri A, Ferrari V, Moglia A, et al. Mixed reality for robotic treatment of a splenic artery aneurysm. Surg Endosc 2010 May 14;24(5):1204. [doi: 10.1007/s00464-009-0703-0] [Medline: 19826869]
40. Mochizuki Y, Hosaka A, Kamiuchi H, Nie JX, Masamune K, Hoshina K, et al. New simple image overlay system using a tablet PC for pinpoint identification of the appropriate site for anastomosis in peripheral arterial reconstruction. Surg Today 2016 Dec 17;46(12):1387-1393. [doi: 10.1007/s00595-016-1326-4] [Medline: 26988854]
41. Mangina E, Almaksy A, Campbell A. 3D modeling for augmented reality systems in novel vascular models. In: Proceedings of the 10th EAI International Conference on Simulation Tools and Techniques. 2017 Presented at: SIMUTOOLS '17: 10th EAI International Conference on Simulation Tools and Techniques; Sep 11 - 13, 2017; Hong Kong China. [doi: 10.1145/3173519.3173523]
42. 42 BE, Felle P, Crowley C, Jones J, Mangina E, Campbell A. Augmented reality EVAR training in mixed reality educational space. IEEE Global Engineering Education Conference, EDUCON. IEEE Computer Society. pp. 1571– 2017:9. [doi: 10.1109/educon.2017.7943058]
43. Bartesaghi S, Colombo G, Morone S. Spatial augmented reality and simulations to improve abdominal aortic aneurysm diagnosis and monitoring. Comput Aided Des Appl 2015 Jun 19;12(6):803-810. [doi: 10.1080/16864360.2015.1033347]
44. Rudarakanchana N, Van Herzeele I, Bicknell CD, Riga CV, Rolls A, Cheshire NJ, et al. Endovascular repair of ruptured abdominal aortic aneurysm: technical and team training in an immersive virtual reality environment. Cardiovasc Intervent Radiol 2014 Aug 7;37(4):920-927. [doi: 10.1007/s00270-013-0765-1] [Medline: 24196270]
45. Anderson J, Chui C, Cai Y, Wang Y, Li Z, Ma X, et al. Virtual reality training in interventional radiology: the Johns Hopkins and Kent Ridge digital laboratory experience. Semin intervent Radiol 2002;19(2):179-186. [doi: 10.1055/s-2002-32796]
46. Sun Q, Mai Y, Yang R, Ji T, Jiang X, Chen X. Fast and accurate online calibration of optical see-through head-mounted display for AR-based surgical navigation using Microsoft HoloLens. Int J Comput Assist Radiol Surg 2020 Nov 18;15(11):1907-1919. [doi: 10.1007/s11548-020-02246-4] [Medline: 32809184]
47. Andrews CM, Henry AB, Soriano IM, Southworth MK, Silva JR. Registration techniques for clinical applications of three-dimensional augmented reality devices. IEEE J Transl Eng Health Med 2021;9:1-14. [doi: 10.1109/jtehm.2020.3045642]
48. Wang X, Habert S, Zu Berge CS, Fallavollita P, Navab N. Inverse visualization concept for RGB-D augmented C-arms. Comput Biol Med 2016 Oct 01;77:135-147. [doi: 10.1016/j.compbiomed.2016.08.008] [Medline: 27544070]

XSL•FO
RenderX

49. Horn M, Ernst F, Kuhlemann I, Stahlberg E, Goltz JP, Wiedner M, et al. Experimental visualization of vascular structures with Microsoft Hololens – set-up for navigated contrast agent and radiation-sparing endovascular procedures (NAV-CARS EVAR). Eur J Vasc Endovasc Surg 2019 Dec;58(6):e104. [doi: 10.1016/j.ejvs.2019.06.640]
50. van den Berg JC. Three-dimensional image overlay to assist endovascular procedures. Vascular Disease Manag 2013;10(9):E179-E184 [FREE Full text]
51. Burke E, Mangina E, Campbell A, Crowley C, Felle A, Jones J. Augmented reality EVAR training in mixed reality educational space. In: Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON). 2017 Presented at: 2017 IEEE Global Engineering Education Conference (EDUCON); Apr 25-28, 2017; Athens, Greece. [doi: 10.1109/educon.2017.7943058]
52. Cheng SWK. Novel endovascular procedures and new developments in aortic surgery. Br J Anaesth 2016 Sep;117 Suppl 2:ii3-ii12 [FREE Full text] [doi: 10.1093/bja/aew222] [Medline: 27566806]
53. Ferrari V, Carbone M, Condino S, Cutolo F. Are augmented reality headsets in surgery a dead end? Expert Rev Med Devices 2019 Dec 22;16(12):999-1001. [doi: 10.1080/17434440.2019.1693891] [Medline: 31725347]
54. Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, et al. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2018;2(1):2 [FREE Full text] [doi: 10.1186/s41747-017-0033-2] [Medline: 29708204]
55. Park BJ, Hunt SJ, Martin C, Nadolski GJ, Wood BJ, Gade TP. Augmented and mixed reality: technologies for enhancing the future of IR. J Vasc Interv Radiol 2020 Jul;31(7):1074-1082 [FREE Full text] [doi: 10.1016/j.jvir.2019.09.020] [Medline: 32061520]

Abbreviations

AR: augmented reality
CBCT: cone-beam computed tomography
CT: computed tomography
DSA: digital subtraction angiography
EVAR: endovascular aneurysm repair
HMD: head-mounted display
MR: mixed reality
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PROSPERO: international prospective register of systematic reviews
TEVAR: thoracic endovascular aneurysm repair
US: ultrasound
VR: virtual reality

Edited by N Zary; submitted 26.10.21; peer-reviewed by F Velayati, R de Leeuw, H Ayatollahi, J Egger; comments to author 16.01.22; revised version received 22.04.22; accepted 23.06.22; published 23.09.22.

Please cite as:
Eves J, Sudarsanam A, Shalhoub J, Amiras D
Augmented Reality in Vascular and Endovascular Surgery: Scoping Review
JMIR Serious Games 2022;10(3):e34501
URL: https://games.jmir.org/2022/3/e34501
doi:10.2196/34501
PMID:36149736

©Joshua Eves, Abhilash Sudarsanam, Joseph Shalhoub, Dimitri Amiras. Originally published in JMIR Serious Games (https://games.jmir.org), 23.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX