# Contents

## Original Papers

### A Safe Place to Learn: Peer Research Qualitative Investigation of gameChange Virtual Reality Therapy (e38065)

### Engagement With Gamification Elements in a Smoking Cessation App and Short-term Smoking Abstinence: Quantitative Assessment (e39975)
Nikita Rajani, Luz Bustamante, Dominik Weth, Lucia Romo, Nikolaos Mastellos, Filippos Filippidis.  

### A Serious Game About Hematology for Health Care Workers (SUPER HEMO): Development and Validation Study (e40350)

### A Software Engineering Framework for Reusable Design of Personalized Serious Games for Health: Development Study (e40054)
Stéphanie Carlier, Vince Naessens, Femke De Backere, Filip De Turck.  

### Vertical Integration of Electronic Health Records in Medical Consortiums: Dynamic Modeling Approach Based on the Evolutionary Game Theory (e41528)
Shenghu Tian, Yu Chen.  

### Games in Times of a Pandemic: Structured Overview of COVID-19 Serious Games (e41766)
Tjasa Kermavnar, Valentin Visch, Pieter Desmet.  

### A Gamified Real-time Video Observed Therapies (GRVOTS) Mobile App via the Modified Nominal Group Technique: Development and Validation Study (e43047)
Siti Abas, Nurhuda Ismail, Yuslina Zakaria, Ismassabah Ismail, Nurul Mat Zain, Siti Yasin, Khalid Ibrahim, Asmah Razali, Mas Mohd Yusoff, Norliza Ahmad, Thilaka Chinnayah.  

### Evaluating the Utility of a Psychoeducational Serious Game (SPARX) in Protecting Inuit Youth From Depression: Pilot Randomized Controlled Trial (e38493)
Yvonne Bohr, Leah Litwin, Jeffrey Hankey, Hugh McCague, Chelsea Singoorie, Mathijs Lucassen, Matthew Shepherd, Jenna Barnhardt.  

### Co-design of a Virtual Reality Cognitive Remediation Program for Depression (bWell-D) With Patient End Users and Clinicians: Qualitative Interview Study Among Patients and Clinicians (e43904)
Maria Hernandez Hernandez, Erin Michalak, Nusrat Choudhury, Mark Hewko, Ivan Torres, Mahesh Menon, Raymond Lam, Trisha Chakrabarty.
Effectiveness of a Personalized, Chess-Based Training Serious Video Game in the Treatment of Adolescents and Young Adults With Attention-Deficit/Hyperactivity Disorder: Randomized Controlled Trial (e39874)
Maria Rodrigo-Yanguas, Marina Martín-Moratinos, Carlos González-Tardón, Fernando Sanchez-Sanchez, Ana Royuela, Marcos Bella-Fernández, Hilario Bisaco-Fontecilla.
126

Serious Game for the Screening of Central Auditory Processing Disorder in School-Age Children: Development and Validation Study (e40284)
Ana-Marta Gabaldón-Pérez, María Dolón-Poza, Martina Eckert, Nuria Máximo-Bocanegra, María-Luisa Martín-Ruiz, Iván Pau De La Cruz.
1

Serious Games Based on Cognitive Bias Modification and Learned Helplessness Paradigms for the Treatment of Depression: Design and Acceptability Study (e37105)
155

Gamifying Cognitive Behavioral Therapy Techniques on Smartphones for Bangkok’s Millennials With Depressive Symptoms: Interdisciplinary Game Development (e41638)
Poe Sriwatanathamma, Veerawat Sirivesmas, Sone Simatrang, Nobonita Bhowmik.
174

Health Promotion in Popular Web-Based Community Games Among Young People: Proposals, Recommendations, and Applications (e39465)
Philippe Martin, Boris Chapoton, Aurélie Bourmaud, Agnès Dumas, Joëlle Kivits, Clara Eyraud, Capucine Dubois, Corinne Alberti, Enora Le Roux.
190

Impact of a Serious Game (#RedPingüino) to Reduce Facial Self-Touches and Prevent Exposure to Pathogens Transmitted via Hands: Quasi-Experimental Intervention (e45600)
Marta Arévalo-Baeza, Alejandro Viuda-Serrano, Carmen Juan-Llamas, Pablo Sotoca-Orgaz, Iván Asín-Izquierdo.
204

Eating Behavior and Satiety With Virtual Reality Meals Compared With Real Meals: Randomized Crossover Study (e44348)
Alkyoni Glympi, Dorothy Odegi, Modjtaba Zandian, Per Södersten, Cecilia Bergh, Billy Langlet.
218

Developing a Game (Inner Dragon) Within a Leading Smartphone App for Smoking Cessation: Design and Feasibility Evaluation Study (e46602)
Justin White, Marie Salem, Séverine Toussaert, J Westmaas, Bethany Raiff, David Crane, Edward Warrender, Courtney Lyles, Lorien Abroms, Johannes Thrul.
230

A Serious Game to Self-Regulate Heart Rate Variability as a Technique to Manage Arousal Level Through Cardiorespiratory Biofeedback: Development and Pilot Evaluation Study (e46351)
Tony Estrella, Carla Alfonso, Juan Ramos-Castro, Altor Alsina, Lluis Capdevila.
247

Exploring Key Factors Influencing Nursing Students’ Cognitive Load and Willingness to Serve Older Adults: Cross-sectional Descriptive Correlational Study (e43203)
Pei-Lun Hsieh, Yu-Rung Wang, Tien-Chi Huang.
264

Design Considerations for an Exergame-Based Training Intervention for Older Adults With Mild Neuropsychiatric Disorder: Qualitative Study Including Focus Groups With Experts and Health Care Professionals and Individual Semistructured In-depth Patient Interviews (e37616)
Patrick Manser, Manuela Adcock-Omlin, Eling de Bruin.
274

Differences in Brain Activity and Body Movements Between Virtual Reality and Offline Exercise: Randomized Crossover Trial (e40421)
Hee Kim, Jea Lee, Gangta Choi, Junghoon Huh, Doug Han.
297
Measured and Perceived Exercise Intensity During the Performance of Single-Task, Cognitive-Motor Dual-Task, and Exergame Training: Transversal Study (e36126)
Matthieu Gallou-Guyot, Anaick Perrochon, Romain Marie, Maxence Bourgeois, Stephane Mandigout. 311

Research Status and Emerging Trends in Virtual Reality Rehabilitation: Bibliometric and Knowledge Graph Study (e41091)
Ting Fan, Xiaobei Wang, Xiaoxi Song, Gang Zhao, Zhichang Zhang. 320

The Effects of Exergaming on Attention in Children With Attention Deficit/Hyperactivity Disorder: Randomized Controlled Trial (e40438)
HongQing Ji, Shanshan Wu, Junyeon Won, Shiyang Weng, Sujin Lee, Sangmin Seo, Jung-Jun Park. 336

The Change in Exergaming From Before to During the COVID-19 Pandemic Among Young Adults: Longitudinal Study (e41553)
Erin O'Loughlin, Catherine Sabiston, Roxy O'Rourke, Mathieu Bélanger, Marie-Pierre Sylvestre, Jennifer O'Loughlin. 349

Usability of the GAIMplank Video Game Controller for People With Mobility Impairments: Observational Study (e38484)
Laurie Malone, Christen Mendonca, Sangeetha Mohanraj, Samuel Misko, Joseph Moore, James Brascorne, Mohanraj Thirumalai. 359

Impact of Gamification on Consumers’ Favorability in Cause-Related Marketing Programs: Between-Subjects Experiments (e35756)
Yanhe Li, Yanchen Li, Xiu Zhou, Kunshu Ma. 376

Finding Effective Adjustment Levels for Upper Limb Exergames: Focus Group Study With Children With Physical Disabilities (e36110)
Martina Eckert, Beatriz Domingo Soria, Noelia Terroso Gil. 386

Negami: An Augmented Reality App for the Treatment of Spatial Neglect After Stroke (e40651)
Britta Stammel, Kathrin Flammer, Thomas Schuster, Marian Lambert, Hans-otto Karnath. 399

Appropriate Image Selection With Virtual Reality in Vestibular Rehabilitation: Cross-sectional Study (e40806)
Kerem Ersin, Emre Gürlek, Hakan Güler, Çi dem Kalaycik Ertugay, Mustafa erbetçio lu. 411

Cognitive Training With Head-Mounted Display Virtual Reality in Neurorehabilitation: Pilot Randomized Controlled Trial (e45816)
Julian Specht, Barbara Stegmann, Hanna Gross, Karsten Krakow. 419

Design Preferences for a Serious Game–Based Cognitive Assessment of Older Adults in Prison: Thematic Analysis (e45467)
Rhys Mantell, Adrienne Withall, Kylie Radford, Michael Kasumovic, Lauren Monds, Ye Hwang. 437

Computerized Block Games for Automated Cognitive Assessment: Development and Evaluation Study (e40931)
Xiangyi Cheng, Grover Gilmore, Alan Lerner, Kiju Lee. 451

Advantages of a Training Course for Surgical Planning in Virtual Reality for Oral and Maxillofacial Surgery: Crossover Study (e40541)
Max Ulbrich, Vincent Van den Bosch, Andrea Bönsch, Lennart Gruber, Mark Ooms, Claire Melchior, Ila Motmaen, Caroline Wipert, Ashkan Rashad, Torsten Kuhlen, Frank Hözle, Behrus Puladi. 469

Design and Evaluation of Using Head-Mounted Virtual Reality for Learning Clinical Procedures: Mixed Methods Study (e46398)
Siew Lau, Rosalind Siah, Khairul Dzakirin Bin Rusli, Wen Loh, John Yap, Emily Ang, Fui Lim, Sok Liew. 484
Effects of Exergaming on Musculoskeletal Pain in Older Adults: Systematic Review and Meta-analysis (e42944)
Nan Mo, Jin Feng, Hai Liu, Xiao Chen, Hui Zhang, Hui Zeng .......................................................... 498

A Conceptual Framework (2D-ME) for Explaining Self–first and Self–third Person Views of Prototyping Dynamics in Serious Games Design: Experimental Case Study (e41824)
Sofia Hadjileontiadou, Sofia Dias, Leontios Hadjileontiadis .......................................................... 521

Role of Gaming Devices Associated With Internet Gaming Disorder in China: Cross-sectional Study (e40130)
Yifan Li, Ying Tang, Shucai Huang, Linxiang Tan, Qiuping Huang, Xinlin Chen, Shuhong Lin, Jingyue Hao, Zhenjiang Liao, Hongxian Shen .......................................................... 538

Virtual Reality Technology in Nursing Professional Skills Training: Bibliometric Analysis (e44766)
Chengang Hong, Liping Wang ........................................................................................................... 624

Gamification and Soft Skills Assessment in the Development of a Serious Game: Design and Feasibility Pilot Study (e45436)
Luca Altomari, Natalia Altomari, Gianpaolo Iazzolino ...................................................................... 655

Preference of Virtual Reality Games in Psychological Pressure and Depression Treatment: Discrete Choice Experiment (e34586)
Shan Jin, Zijian Tan, Taoran Liu, Sze Chan, Jie Sheng, Tak-hap Wong, Jian Huang, Casper Zhang, Wai-Kit Ming .................................................................................................................. 671

Viewpoint

Time to Think “Meta”: A Critical Viewpoint on the Risks and Benefits of Virtual Worlds for Mental Health (e43388)
Vincent Paquin, Manuela Ferrari, Harmehr Sekhon, Soham Rej .......................................................... 512

Reviews

Mixed Reality in Modern Surgical and Interventional Practice: Narrative Review of the Literature (e41297)
Mats Vervoorn, Maaike Wulfse, Tristan Van Doormaal, Jelle Ruurda, Niels Van der Kaaij, Linda De Heer ......................................................................................................................... 549

Children and Young People’s Involvement in Designing Applied Games: Scoping Review (e42680)
Michael Saiger, Sebastian Deterding, Lina Gega ................................................................................. 560

Effectiveness of Digital Health Interventions Containing Game Components for the Self-management of Type 2 Diabetes: Systematic Review (e44132)
Linda Ossenbrink, Tina Haase, Patrick Timpel, Olaf Schoffer, Madlen Scheibe, Jochen Schmitt, Stefanie Deckert, Lorenz Harst .......................................................................................................... 584

Comparison of Exergames Versus Conventional Exercises on the Health Benefits of Older Adults: Systematic Review With Meta-Analysis of Randomized Controlled Trials (e42374)
Xi Chen, Lina Wu, Hui Feng, Hongling Ning, Shuang Wu, Mingyue Hu, Dian Jiang, Yifei Chen, Yu Jiang, Xin Liu .......................................................................................................................... 611

Dimensions of Interactive Pervasive Game Design: Systematic Review (e42878)
Liu Kai, Wee Tan, Erni Saari ................................................................................................................ 639
A Safe Place to Learn: Peer Research Qualitative Investigation of gameChange Virtual Reality Therapy

Jessica Bond, MSc; Alexandra Kenny, BA; Vanessa Pinfold, PhD; Lisa Coperthwaite, BA; The gameChange Lived Experience Advisory Panel; Thomas Kabir, PhD; Michael Larkin, PhD; Ariane Beckley, MA, MSc; Laina Rosebrock, PhD; Sinéad Lambe, DClinPsy; Daniel Freeman, DClinPsy; Felicity Waite, DClinPsy; Dan Robotham, PhD

1McPin Foundation, London, United Kingdom
2School of Psychology, Aston University, Birmingham, United Kingdom
3Department of Psychiatry, University of Oxford, Oxford, United Kingdom
4Oxford Health National Health Service Foundation Trust, Oxford, United Kingdom
5National Institute for Health and Care Research Oxford Health Biomedical Research Centre, Oxford, United Kingdom

Corresponding Author:
Jessica Bond, MSc
McPin Foundation
7-14 Great Dover Street
London, SE1 4YR
United Kingdom
Phone: 44 207 922 7877
Email: contact@mcpin.org

Abstract

Background: Automated virtual reality (VR) therapy has the potential to substantially increase access to evidence-based psychological treatments. The results of a multicenter randomized controlled trial showed that gameChange VR cognitive therapy reduces the agoraphobic avoidance of people diagnosed with psychosis, especially for those with severe avoidance.

Objective: We set out to use a peer research approach to explore participants’ experiences with gameChange VR therapy. This in-depth experiential exploration of user experience may inform the implementation in clinical services and future VR therapy development.

Methods: Peer-led semistructured remote interviews were conducted with 20 people with a diagnosis of psychosis who had received gameChange as part of the clinical trial (ISRCTN17308399). Data were analyzed using interpretative phenomenological analysis and template analyses. A multiperspectival approach was taken to explore subgroups. Credibility checks were conducted with the study Lived Experience Advisory Panel.

Results: Participants reported the substantial impact of anxious avoidance on their lives before the VR intervention, leaving some of them housebound and isolated. Those who were struggling the most with agoraphobic avoidance expressed the most appreciation for, and gains from, the gameChange therapy. The VR scenarios provided “a place to practise.” Immersion within the VR scenarios triggered anxiety, yet participants were able to observe this and respond in different ways than usual. The “security of knowing the VR scenarios are not real” created a safe place to learn about fears. The “balance of safety and anxiety” could be calibrated to the individual. The new learning made in VR was “taken into the real world” through practice and distilling key messages with support from the delivery staff member.

Conclusions: Automated VR can provide a therapeutic simulation that allows people diagnosed with psychosis to learn and embed new ways of responding to the situations that challenge them. An important process in anxiety reduction is enabling the presentation of stimuli that induce the original anxious fears yet allow for learning of safety. In gameChange, the interaction of anxiety and safety could be calibrated to provide a safe place to learn about fears and build confidence. This navigation of therapeutic learning can be successfully managed by patients themselves in an automated therapy, with staff support, that provides users with personalized control. The clinical improvements for people with severe anxious avoidance, the positive experience of VR, and the maintenance of a sense of control are likely to facilitate implementation.

(JMIR Serious Games 2023;11:e38065) doi:10.2196/38065
KEYWORDS
peer research; lived experience; qualitative methods; interpretative phenomenological analysis; IPA; anxiety; psychosis; virtual reality; VR; cognitive therapy; automated; implementation

Introduction

Background

gameChange is an automated virtual reality (VR) cognitive therapy designed to reduce agoraphobic avoidance and distress in everyday situations. In a multicenter randomized controlled trial with 346 people with psychosis, we found that gameChange led to significant reductions in avoidance and distress [1]. Overall, there were small effect size improvements in agoraphobic avoidance, agoraphobic distress, and perceptions of recovery. However, patients with the most severe agoraphobia, for example, those struggling to leave the home, had large effect size improvements in agoraphobic distress and avoidance as well as significant improvements in persecutory ideation, quality of life, and perceived recovery, which were maintained at 6 months.

Agoraphobic avoidance is understood from a cognitive perspective in gameChange. The avoidance is a defense (or safety-seeking) behavior in response to fearful thoughts [2,3]. Fearful thoughts often develop because of past unpleasant or threatening experiences. However, the fears are maintained as the absence of harm in the present moment is attributed to the use of defenses rather than there being no current threat. Even when in the anxiety-provoking situations, the use of more subtle defenses (eg, rushing or avoiding eye contact) plays the same maintaining role for catastrophic cognitions [4]. In treatment, defenses must be dropped so that the threat beliefs can be evaluated fully. This is not to overlook the fact that many patients with psychosis have had difficult experiences and may currently be living in adverse circumstances and potentially facing ongoing threats. Although there may not be current overt actions against the person, there can be more implicit and subtle forms of threat, particularly linked to stigma. Therefore, clinicians working with patients should always be aware of the potential for actual threat and assess appropriately.

On the basis of cognitive psychology principles, gameChange VR therapy was developed through a user-centered design process involving >500 hours of input from people with lived experience [5,6]. In 6 sessions, people enter interactive computer-generated simulations of everyday scenarios such as a café, a bus, or a physician’s waiting room. Guided by a virtual coach, patients are encouraged to experiment with dropping defense behaviors and to evaluate their fears. A member of the staff is present to assist with the equipment and help plan homework tasks to consolidate the learning from VR. Consistent with the cognitive model, we found in the trial that improvements in agoraphobic anxiety and avoidance were mediated by reductions in both a wide range of threat cognitions and the use of within-situation defense behaviors [1].

Objectives

In this study, we used a peer research approach to try to understand the participant experience of gameChange VR therapy in depth. Peer research is steered and conducted by people with relevant lived experiences. Peer methods have the potential to facilitate greater depth and more nuanced data collection and analysis by enhancing rapport [7] and leveling power [8]. Working collaboratively with researchers from different backgrounds, including lived experience, can enhance the validity and trustworthiness of qualitative analysis [9,10]. Interpretative phenomenological analysis (IPA) is a qualitative approach that focuses on participants’ lived experiences and how they make sense of them [11,12]. IPA was chosen as it explicitly acknowledges the role of the researcher in the interpretation, which complements the peer research approach of valuing the insights that lived experience can bring [11]. Template analysis can be used to extend and develop IPAs for larger samples [13]. We conducted a peer-methods qualitative study incorporating IPA and template analysis to explore participants’ experiences with gameChange.

Methods

Overview

This was a qualitative peer-methods study investigating the experiences with gameChange VR therapy for agoraphobic avoidance. A multiperspectival design was used [14], and a peer research approach was taken, including collaboration with the gameChange Lived Experience Advisory Panel (LEAP). The LEAP was a panel of 12 people with lived experience of psychosis. Full details of the study methodology are provided in the published study protocol [15]. Criteria to ensure credibility in qualitative research [16,17] were used to inform the design and ensure rigor in the conduct of the study.

Ethics Approval

The study received Health Research Authority approval and Health and Care Research Wales approval (IRAS 256895), as well as ethics approval from the National Health Service (NHS) South Central - Oxford B Research Ethics Committee (19/SC/0075) as part of the gameChange trial.

Research Team

People with lived experience of psychosis and anxious social avoidance were involved in all stages of the study. The research team included peer researchers who had had similar experiences to those being explored, a qualitative researcher with lived experience of mental health issues that did not directly overlap with those of the participants, clinical psychologists who designed and used VR therapy, and a qualitative researcher who pioneered the IPA approach. The LEAP, facilitated by the McPin Foundation, a charity that promotes the use of lived experience in research, contributed to the design, conduct, and analysis of the study. We aimed to use the peer knowledge within the study at all stages, including the design, data collection, analysis, and presentation of the findings.
Participants

We recruited participants who had received gameChange VR therapy as part of the trial (ISRCTN17308399). The participants were people accessing NHS mental health services with a diagnosis of psychosis and self-reported difficulties going into everyday social situations because of anxiety. The inclusion and exclusion criteria for the trial are listed in the trial protocol [18]. The additional inclusion criteria for this qualitative study were willingness and ability to provide informed consent to participate in the interview and willingness to have the interview audio recorded.

Sampling and Recruitment

Consistent with IPA principles, the recruitment procedure sought homogeneity—the locus of homogeneity in this study was receipt of gameChange therapy. The primary sampling method was to invite a consecutive cohort of people in the final phase of the trial between April 2021 and September 2021. To explore different experiences of treatment using a multiperspectival design, we sought a larger sample size (N=20), including participants with low, medium, and high uptake of therapy. When possible, this was supplemented with purposive sampling to ensure variation across demographic characteristics (including age, gender, and ethnicity) and trial centers. The recruitment strategy was revised because of the high uptake of therapy in the trial (just under 90% of participants received a dose of therapy, which was defined as a minimum of 3 sessions). Therefore, the subgroups for the multiperspectival design were derived retrospectively from the data and are presented in the Results section.

A total of 27 people were invited to participate, of whom 7 (26%) declined and 20 (74%) consented. Participants were from each of the 5 trial centers that recruited participants from NHS mental health trusts across England: Avon and Wiltshire Mental Health Partnership NHS Trust; Greater Manchester Mental Health NHS Foundation Trust; Cumbria, Northumberland, Tyne, and Wear NHS Foundation Trust; Nottinghamshire Healthcare NHS Foundation Trust; and Oxford Health NHS Foundation Trust. Recruitment was facilitated by the local trial coordinator. Written or audio informed consent, including consent to the use of pseudonymized quotes, was obtained from all participants.

Procedure

The semistructured interview guide (Multimedia Appendix 1) was developed in collaboration with the LEAP, with reference to existing literature, and in line with the principles of IPA [19-21]. The guide was refined following pilot interviews with members of the LEAP. The guide was used flexibly, with a focus on eliciting the participants’ own account. This involved using open-ended descriptive and narrative questions to explore the context, followed by individually tailored prompts to elicit further information or clarification and using the participants’ own language whenever possible. The focus of the interview guide was on participants’ experience with the gameChange therapy. It included questions such as What is your experience of going out of the house or going into social situations? What was it like to be in everyday situations in virtual reality? and Compared to what you told us about your life before gameChange, is life different in any way now? How so?

Interviews were conducted between April 2021 and September 2021. The first 40% (8/20) of interviews were conducted jointly by JB and AK, a further 55% (11/20) were facilitated by AK alone, and 5% (1/20) were facilitated by JB alone. Interviews were conducted remotely, either by phone (14/20, 70%) or web-based video call (6/20, 30%). The mean duration was 80 (SD 22; range 40-105) minutes, with most (13/20, 65%) completed over 2 sessions. The peer researcher chose when, how, and what to disclose about their related mental health experiences, and this differed between interviews. Both interviewers kept field notes, including on the use and impact of sharing the peer identity. Interviews were audio recorded, transcribed verbatim, and deidentified.

Analysis

A total of 25% (5/20) of the transcripts were coded using IPA [11,20]. IPA focuses on participants’ lived experiences and how they make sense of them [11]. The remaining transcripts were analyzed using the initial IPA themes as a preliminary template. This work was conducted using template analysis [22], which can be used with IPA to analyze larger samples [13].

The multiperspectival design provides a structure for exploring both the individual account and related groups [14]. In our study, the subsamples were identified via preliminary analysis rather than via sampling strategies. Four subsamples related to overall experience and the impact of therapy were identified: (1) Big changes (4/20, 20%); (2) Better ways to cope (9/20, 45%); (3) Hard to say or hard to hang on to (5/20, 25%); and (4) Struggling on (2/20, 10%).

The 25% (5/20) of transcripts included in the IPA analysis were selected based on the richness of the data and to ensure a range of trial sites and ages and a balance of gender. To maintain the homogeneity of the sample, a requirement of IPA, the 5 transcripts expressed a similar perspective on gameChange. The analytical procedure for IPA followed the approach outlined by Smith et al [11,20]. This included line-by-line annotation of each transcript, mapping what mattered to each participant (“objects of concern”) with the meaning assigned to the experience (“experiential statements”). Meanings were clustered and interpreted for each case. Across cases, themes were identified that encompassed the phenomenological experiences and understanding of the 25% (5/20) of participants, forming superordinate and subordinate themes. IPA provided the foundation for the subsequent template analysis.

A provisional template was developed from the IPA themes and used to initiate the template analysis of the remaining data. The template analysis followed the steps outlined by King et al [22]. In line with the multiperspectival design [14], the transcripts were introduced in subgroups, focusing on each in turn. Analyzing the data in these batches made comparisons between perspectives starker. New codes were assigned and the template was updated at least once after each group was added. Once all data had been analyzed, the template was reviewed against all transcripts. Two analysis sessions were held with the LEAP to develop the template. LEAP members were given an
Overview of the themes and shown a selection of quotes. The focus was on identifying patterns within the data, resolving ambiguity in the accounts, and considering how the participants’ experiences resonated with their own.

The IPA analysis was conducted by JB, AK, and LC. DR and VP contributed to the template analysis. Supervisory support was provided at each stage of the analysis by ML and FW.

Reflexivity and Credibility
All members of the research team reflected on the different perspectives that they brought to the study design, setup, data collection, and analysis. For example, in the analysis, the peer researchers drew on their lived experience to support their interpretation of meaning in the data; this was particularly relevant when exploring meanings related to participants’ experiences of anxious avoidance and psychosis.

Credibility checks were conducted with members of the gameChange LEAP. Their input was used to corroborate and deepen the interpretation of the themes. The reflexive logs, transcripts, interview schedule development, and minutes of supervisory research team meetings provided an audit trail for analysis. Illustrative quotes, description of the study context, and participant characteristics allowed for an assessment of the transferability of the findings to other settings and contexts.

Results
Contextualizing the Data
A total of 20 people participated: 11 (55%) women and 9 (45%) men. The average age of the participants was 37 (SD 12.9) years. In total, 85% (17/20) of the participants identified as White British individuals, 10% (2/20) identified as British Asian individuals, and 5% (1/20) identified as Polish. As can be seen in Table 1, most (15/20, 75%) participants were unemployed. A total of 20% (4/20) of the participants said that they spent a lot of time playing computer games and described themselves as “gamers.” In total, 60% (12/20) of the participants attended 6 sessions of therapy, and 40% (8/20) attended 7 sessions. For 55% (11/20) of the participants, these VR sessions were facilitated by a clinical psychologist; for 30% (6/20) of the participants, it was an assistant psychologist; and 15% (3/20) of the participants worked with peer support workers. The participant characteristics in this qualitative study were consistent with those in the main trial (see the trial outcome report [1]).

Table 1. Participant characteristics (N=20).

<table>
<thead>
<tr>
<th>Demographic and clinical characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Married</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Cohabiting</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Retired</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Carer</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Employed (full-time)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Student (part-time)</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Living situation</strong></td>
<td></td>
</tr>
<tr>
<td>Living with parents or another relative</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Living alone</td>
<td>7 (35)</td>
</tr>
<tr>
<td>Living with others</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Living with a spouse or partner</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Mental health service context at the time of interview</strong></td>
<td></td>
</tr>
<tr>
<td>CMHT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15 (75)</td>
</tr>
<tr>
<td>EIP&lt;sup&gt;b&lt;/sup&gt; team</td>
<td>5 (25)</td>
</tr>
</tbody>
</table>

<sup>a</sup>CMHT: community mental health team.
<sup>b</sup>EIP: early intervention in psychosis.

Overview of Findings
An overview of the key findings is presented. This is followed by a description of the subgroups identified for the multiperspectival design. The cross-case analysis of participants’ experiences with the intervention is then presented. Illustrative quotes, with details omitted to protect anonymity, are presented in the text and in Multimedia Appendix 2.
Summary of Key Findings

There were 5 superordinate themes (Textbox 1). First, anxious avoidance carries a cost. Participants described living restricted lives in which fears made it hard to do everyday activities. Second, participants shared their curiosity and motivation to try VR therapy. Third, VR therapy provided a place to learn about anxiety and practice different or new responses. Fourth, the VR simulations triggered an anxiety response, yet participants simultaneously recognized that the environments were largely safe. This interaction of anxiety and safety could be calibrated by the individual to provide a safe place to learn about fears and build confidence. Finally, VR was a training ground to do things differently. Through practice and conversations with the member of the staff who was present, participants took the key learning from VR into the real-world situations that mattered to them.

Textbox 1. Overview of superordinate and subordinate themes.

- Experience and cost of anxious avoidance
- Reasons to try: curiosity and motivation
- A place to practice
  - An immersive experience
  - A chance to observe anxiety
  - New ways of responding
- The security of knowing it is not real
  - The sweet spot of safety and anxiety
  - Calibrating for a personalized approach
- Taking it into the real world
  - From training wheels to real-world practice
  - One thing to hold onto

The 4 Perspectives

Overview of the 4 Perspectives

We identified 4 subgroups of participants based on the descriptions of their experience and impact of the gameChange therapy. The descriptions varied from life-changing to experiencing little change, although all described gameChange as a valuable experience. Illustrative quotes for each subgroup are presented in Table 2.
Table 2. Illustrative quotes from the participant subgroups (N=20).

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Participants, n (%)</th>
<th>Illustrative quotes</th>
</tr>
</thead>
</table>
| Big changes         | 4 (20)              | “It’s very different now, I can open my curtains and look out my window, I can go out to local shops and places, I can meet up with friends and family, I can do such a lot more.” [Participant 27]
|                     |                     | “Before I would never do that, ever. But I did that and I was talking to new people and I was having a great time. It’s just being in social situations instead of hiding from them, I’m just trying to embrace them.” [Participant 6] |
| Better ways to cope | 9 (45)              | “It definitely helped with anxiety, I still suffer from some of the things I did, well, all of the things I did before, but it’s definitely, definitely helped it in some way.” [Participant 3]
|                     |                     | “Before I’m just looking at the floor wherever I’m going but now I’m looking at everyone, like, in their eyes and stuff. Even though I’m still too scared to go to the supermarket on my own, I have been able to walk past it on my own, like walk up there and past it, whereas before I wouldn’t have been able to do that. I’ve been in small shops on my own and the bus, I can do that on my own now without worrying.” [Participant 25] |
| Hard to say or hard to hang onto | 5 (25) | “Now because my mood is quite good, I’m thinking, ‘Yeah, actually. I did take away stuff.’ But when I’m really anxious, like I have been really anxious over the last couple of weeks...and then I’m like, I haven’t learnt anything.” [Participant 17]
|                     |                     | “I mean I wasn’t too bad with a lot of the situations anyway, so I don’t know if I’ve taken a lot from it, but I’d like to think I have but I’m not entirely sure. I’m not entirely sure if it’s had any massively notable impact on me.” [Participant 13] |
| Struggling on       | 2 (10)              | “It has not really made a magical difference for me, unfortunately, but I can see how it can be very helpful. I think I’m just too entrenched in my avoidance little bubble, that it wasn’t strong enough to pull me out of it. It’s just the same.” [Participant 19] |
|                     |                     | “Like, I say, still go to cafes, still get on the bus, go shopping. I didn’t go food shopping yesterday. I was really, really anxious yesterday in town with the boys. I couldn’t do my food shopping. My body wouldn’t let me do it. My mind affects my body.” [Participant 26] |

**Big Changes**

A total of 20% (4/20) of the participants described gameChange as a transformative experience. Participants in this group marveled at how far they had come—they were able to do things they had not imagined at the start of therapy, when they were often housebound and experiencing very low mood. These accounts were characterized by excitement and enthusiasm.

**Better Ways to Cope**

A total of 45% (9/20) of the participants described experimenting with new ways to approach and cope with situations despite some ongoing anxiety. For these participants, social situations were less stressful but not yet comfortable.

**Hard to Say or Hard to Hang Onto**

A total of 25% (5/20) of the participants noted some small changes following therapy but found these hard to identify when asked directly or hard to maintain when anxiety increased or mood dipped. Participants valued aspects of the therapy but did not necessarily find that the VR scenarios triggered their anxiety. Participants in this group were often able to engage in some social situations before starting therapy.

**Struggling On**

In total, 10% (2/20) of the participants recognized the potential of the therapy but described little personal benefit as they struggled with psychotic experiences. They continued to feel anxious in the familiar and often essential (eg, shopping) situations they were able to engage in even before the start of therapy.

**Main Themes**

**The Experience and Cost of Anxious Avoidance**

All participants spoke about the cost of anxious avoidance. Some participants described themselves as almost housebound, others avoided specific situations, and a few could perform essential tasks in an area where they felt safe. However, all felt the burden and restriction of anxious avoidance:

> It affects my life big time because I can’t do half the things that I want to do or what a normal everyday person would do. I want to do something but sometimes it can hold me back and it can really affect the way you think, the way you act and the person that you are. [Participant 27, Big changes subgroup]

As in the previous extract, participants shared the reasons for withdrawal: many felt threatened by others, at risk of physical or social harm, or that others were observing or talking about them. Some were anxious about showing symptoms in public as they feared that this would draw attention and invite judgment. Participants described the tactics that they had developed to try to reduce these risks, but some strategies had life-limiting effects. Some described their situation as intolerable, affecting self-esteem and quality of life, disrupting careers, impeding friendships, and making them feel “almost not like a human being.” Although all participants described life being “limited” by anxious avoidance, not all were distressed by inhabiting a small world.

**Reasons to Try: Curiosity and Motivation**

As VR therapy was new to the participants, most had no preconceived notions of what gameChange would be like or
how it would help. Many said that they were “open-minded” to try gameChange. However, the reasons to engage differed:

I went in with an open mind. I was willing to give anything a go because I just didn’t want to be in the same state I was in after episode one. [Participant 3, Better ways to cope subgroup]

I’m interested in technology, new technology. It’s quite exciting. So, that really encouraged me to get involved as well. [Participant 1, Big changes subgroup]

As mentioned previously, for some, the opportunity appeared when they were ready to receive and engage with it. Others noted that experiencing a difficult time before gameChange fueled their desire for change. This was often accompanied by a sense of personal responsibility for the intervention’s success. Some were open to gameChange as they had been unwell for a long time and were keen to try anything that could alleviate their symptoms. Others simply saw it as an opportunity to do something in an otherwise restricted life: “I don’t really have anything to do so I was like: ‘Go on then, I might as well do it’” (participant 7, Hard to say subgroup).

The use of VR in gameChange was appealing to more than half (11/20, 55%) of the participants. Those who had heard of VR were attracted by its novelty, they were “interested” or “intrigued” to see what it would be like. They understood its therapeutic use as “a modern new way to treat social anxiety”:

I was, like, curious, I suppose, yeah. Its cutting-edge technology, isn’t it? [Participant 26, Struggling on subgroup]

Of the 4 participants who identified as gamers, 3 (75%) were primarily attracted to the intervention as it was a chance to use VR.

A Place to Practice

I really enjoyed it. Well, I say enjoyed it, it was actually quite hard work. I found it very realistic when I went in, more than I expected, I think. And it was a lot harder than I thought it was going to be. [Participant 1, Big changes subgroup]

Most (14/20, 70%) participants, including one of the 4 gamer participants, were surprised by how realistic it was. Participants seemed to be struck by its immersive quality:

Going into VR was like going into another world, kind of thing. It was really like everything was not realistic, you could tell it was all fake, but at the same time you felt like you were actually in there. [Participant 25, Better ways to cope subgroup]

I was like being in a film, it was. You put the helmet [VR headset] on and you’re taken to another world. [Participant 26, Struggling on subgroup]

Immersion was created by the graphics, background noise, and interactive details:

You can open the door. You can lift a can of beans up and put your bank card on the meter to pay for it.

I thought it was amazing. [Participant 21, Better ways to cope subgroup]

This immersion and perceived realism meant that many participants saw the VR scenarios as a meaningful environment to practice in. Participants who identified as gamers tended to be less impressed with the VR world.

Participants described developing new insights into their experience of anxiety, as described in the following extract:

I didn’t realise things like I was looking for an escape route and going to it. And now, I’m aware of that, I cannot do it so much. You know, it’s just sort of being aware. [Participant 1, Big changes subgroup]

Being in the VR simulations could trigger anxiety. Participants spoke of sweaty palms and racing hearts and of realizing that they were unconsciously trying to use coping mechanisms such as looking for an escape route or exit, avoiding eye contact, or diminishing their physical presence. For some, this was a hard and shocking realization:

...that was really an eye-opener, and I was quite shocked at how it has affected my life for so long. [Participant 27, Big changes subgroup]

VR provided an opportunity to observe anxiety and automatic responses with some emotional distance. It was “a buffer” or a chance “to step back and think about how they are reacting,” as members of the LEAP described it. Responses that were overwhelming and panicked in real life could be slowed, picked apart, and rerun. A LEAP member commented that VR was like a magnifying glass, clarifying the details and nuances of the situation. This led to participants developing greater understanding of what happens in moments of anxiety. This could be empowering and raised the potential to change these automatic responses.

Participants described VR as a place to practice new ways of responding in situations:

Make eye contact. I smile a bit more at people. Little things you don’t realise you haven’t been doing till you go and do the headset. It made me realise a lot, that I had just focused my life around the voices and not taken any notice of the outside world. That’s why I’ve been taking more notice of things when I’m out now. [Participant 23, Better ways to cope subgroup]

The automated virtual coach and staff member made suggestions to drop defenses and try new strategies to build confidence. Despite the anxiety, it was safe to try new approaches as participants knew that the VR was not real.

“The Security of Knowing It Is Not Real”

The nature of VR means that, despite knowing that it is a simulation, reactions are typically real—as evidenced in the anxiety responses that the participants experienced:

Even though you’d tell yourself that you’re in a safe environment and the people weren’t going to necessarily react as you would expect a real-life person to react, your brain is still telling you that you’re in danger and you’re at risk here so yes, it did

https://games.jmir.org/2023/1/e38065
make you feel quite anxious. [Participant 2, Big changes subgroup]

However, knowing that the VR simulations were not real allowed participants to feel a sense of safety:

It makes you face your fears but in a less extreme way because these people, you know aren’t real, but you still have the feeling of it being people. [Participant 16, Better ways to cope subgroup]

Participants described an active interaction between safety and anxiety. For example, a sense of safety was required for participants to “slowly push themselves out of their comfort zone” (participant 8, Big changes subgroup), try the new ways of responding that the automated virtual coach and staff member suggested, and learn to tolerate discomfort. This was a delicate balance. If anxiety was “dialed” too low, VR was perceived as boring and ineffective. If safety was “dialed” too low (or anxiety too high), VR was intense and draining:

It’s a fine line: you want it to be realistic enough, so it’s a good practice, but it’s also quite helpful if it’s not totally realistic because then you get that security of knowing that it’s not real. [Participant 1, Big changes subgroup]

As mentioned previously, participants understood the importance of locating this “sweet spot,” noting that it was sometimes helpful that the VR was not photorealistic or that they knew they were in no danger despite feeling bodily symptoms of anxiety. Finding the right balance could be difficult. For some participants, aspects of the VR scenarios matched real-world triggers, leading to fears or memories of past difficulties and resulting in spikes in anxiety. At these moments, it was sometimes hard to feel safe enough to create new learning in VR.

Levels of anxiety and safety could be calibrated by and to the individual through choice of the levels within the program or by the actions of the staff member:

The deliverer knew if I was putting defences up and she would tell me different ways to cope. There was one scenario where I was in the street and people walked past and I put my defences up because I kept looking at them walking past and my therapist said to relax and look at the cars and describe what’s in the cars and stuff. [Participant 11, Hard to say or hard to hang onto subgroup]

The features of the software that often increased anxiety were “realness” and immersion in the scenarios, background noise, the scenarios becoming more crowded, and specific tasks such as waiting in a queue. Features that increased the sense of safety were that the characters and scenarios did not look as real as the real-world situations, the gradual increase in intensity of the levels, and having the option of removing the headset.

There were layers of support from the virtual coach, Nic, and the staff member. Those with higher baseline levels of anxiety described the coach as comforting, and participants appreciated her guidance and encouragement:

When I did a level, she would be like, “Well done, you did great.” She was good in that sense. It was someone familiar that I would see every time I went in. [Participant 25, Better ways to cope subgroup]

Those who were more confident in VR described the coach as “repetitive,” “slow,” and at times “irritating.”

The staff member was able to build on the generic support that was programmed in the software through the virtual coach, increasing the participants’ sense of safety to facilitate more exploration and experimentation in VR. The staff member was attuned to the participants’ responses, suggesting pauses, encouraging repetition of scenarios or levels that they found challenging, or helping identify and drop specific defenses:

I think the deliverer was quite helpful in saying: “In that one you appeared more anxious, do you want to do that one again? You spoke a bit quietly in that one, did you want to do it again, you didn’t seem that confident doing that one, would you like to try it again?” [Participant 15, Better ways to cope subgroup]

I would use coping strategies. I didn’t realise until the deliverer would tell me what I did—I didn’t recognise it myself—so, the deliverer would tell me certain things that would give me more confidence for the VR, but also, when leaving the session as well. [Participant 3, Better ways to cope subgroup]

In this way, the sweet spot of anxiety and safety could usually be found for each person, creating the conditions for them to try new ways of responding in the VR scenario.

Taking It Into the Real World

The fear was there but what I was fearing didn’t actually happen which was good. That, for me, was part of the therapy in the sense that it made you realise that what you fear is going to happen isn’t necessarily what will happen. It made you re-evaluate if you like your way of thinking. [Participant 2, Big changes subgroup]

Participants understood that VR was a safe place to do things that scared them—“training wheels” where they could take “baby steps”—before translating them to the real world:

The everyday situations that I found difficult, the more I practised them in the VR, the more I could get confident and be more confident in day-to-day life. I think because I was learning about it in the VR and practising and practising and practising, I could then take that and build up more confidence and do it in the everyday real world. [Participant 27, Big changes subgroup]

I guess it’s like the training wheels, isn’t it? You get used to a situation simulating it and then go and try it in the real world. [Participant 5, Hard to say or hard to hang onto subgroup]

Many participants recognized the need to practice between VR sessions and appreciated being given “homework.” Some acknowledged that this structure helped them push themselves.
Others found additional ways to reinforce learning, such as writing notes and asking questions, periods of active reflection, and debriefing with the deliverer and other people. Some used sessions with their care teams to implement homework tasks. These participants were not deterred by circumstances such as COVID-19 restrictions hindering what they could do in the real world, and they were able to adjust the tasks around them. These participants tended to describe gaining the most benefit from the therapy and were in the Big changes and Better ways to cope subgroups.

Participants needed motivation to face anxiety-provoking situations—the potential rewards needed to be high enough for the effortful process of overcoming anxiety. This was observed in the descriptions of the 15% (3/20) of participants who already felt safe in their local environment and were unsure about the value of going further afield. These participants were in the Struggling on and Hard to say subgroups. For some, the relationship with the staff member was an important motivator to put into practice the learning from VR.

Through practicing in VR, participants described creating new learning and finding new explanations and understandings about themselves, other people, and the world. A common realization was that reality was not as bad as participants feared. The seed for these new beliefs was often found in VR—opportunities to make this discovery were embedded in the VR scenarios, for example, when characters smiled during interactions or simply did not respond as feared (eg, laughing or criticizing the participant). This new learning was tailored and reinforced by the staff member, ensuring that the key messages were personalized, succinct, and easy to recall. This learning often became most powerful when participants grew confident enough to push themselves in real life, making eye contact or small talk with people. These layers of corroboration, including discovery in VR, discussion with the deliverer, and real-world practice, made for a powerful learning experience. As well as embedding new beliefs, participants held onto memories created during gameChange. Some spoke of recalling scenes in VR or of ways of thinking and behaving. Therapeutic gains arose out of techniques to gain distance on anxious responses and try new alternative ways of thinking and behaving. Therapeutic gains arose out of a delicate balance of experiencing typical anxious responses but only to a degree in which safety could also be kept in mind. This had a calming effect when they found themselves in a challenging situation:

*It’s like [the staff member] says, “stand confident and people will see you as confident. No one is looking at you. People do look, it doesn’t say they’re going to hurt you.” It really helped me, because when I went out I thought of the gameChange.* [Participant 21, Better ways to cope subgroup]

The result of this practice and learning was that some participants saw changes that were important to them. The world had opened up for people whose lives were most restricted by agoraphobic avoidance:

*It’s a case of not thinking twice about getting the bus and getting take away coffee now. I’m quite happy to do it which was a big change for me.* [Participant 2, Big changes subgroup]

**Discussion**

**Principal Findings**

Facilitated by peer methods, participants provided rich and informative accounts of the problem of agoraphobic anxiety and avoidance, the experience with automated VR therapy, and its impact on their lives. The restriction caused by severe agoraphobic avoidance has a substantial cost in people’s lives. Variations in the benefits of VR therapy were observed across the participant groups. Participants who were close to housebound (4/20, 20%) tended to report “big changes” from the gameChange VR therapy, whereas individuals who were reasonably comfortable with the boundaries of the locations that they could visit (5/20, 25%) tended not to report so much impact from the VR sessions. This perspective from the qualitative interviews of variation in benefits of VR therapy was gained before the trial outcome results were known by those involved in the qualitative analysis but were entirely consistent with the main trial findings. The largest group identified (9/20, 45%) derived benefit from gameChange—they developed “better ways to cope”—but in a less transformational way than the group who saw “big changes.” Their accounts suggest that, before the intervention, they were more able to perform everyday activities than the other group, but they were still distressed by their agoraphobia. These results support the conclusion that the implementation of gameChange is best focused first on people with the most severe agoraphobic difficulties.

There was further consistency with the quantitative findings. Participants reported a wide range of fears contributing to agoraphobic avoidance, as has been found using self-report assessments [4]. Furthermore, the participants viewed VR as a popular treatment option with limited mention of any side effects, which is consistent with our surveys across the whole group that received gameChange [23]. However, the interviews provided greater depth on several important issues. They provided an intriguing perspective on the therapeutic nature of immersive technology. Participants reported how VR—as they knew they were experiencing simulations—provided the chance to gain distance on anxious responses and try new alternative ways of thinking and behaving. Therapeutic gains arose out of a delicate balance of experiencing typical anxious responses but only to a degree in which safety could also be kept in mind. In this way, safety feelings could begin to feel like a realistic appraisal of the situations. This calibration was largely determined by the users, for example, selecting the scenario or level within the program, but there could also be invaluable support from the staff member who was present. A degree of guided support, even with automated therapies, is likely to be very important. It was also clear that participants considered that, to make change, it was important to apply the learning to real-world situations via the homework tasks. Conversations with the supporting staff member were valued. A common realization among those who gained benefits was that reality
was not as bad as they feared. There are opportunities to further tailor the treatment to the individual by increasing the range of scenarios available and adapting the demographics of the virtual coach.

Conclusions
The interviews provided fascinating in-depth insights into the experience with automated VR therapy, which will inform the delivery of gameChange alongside other implementation and health economic studies [23-25]. This study has several limitations. The participants interviewed were not representative of all those who received VR therapy. For instance, we only succeeded in interviewing participants who completed therapy, which is largely a reflection of the high uptake of VR therapy in the trial. This means that we would have missed important specific challenges and barriers for individuals who did not complete VR therapy. The participant group was also not representative of the wider population of people attending mental health services. There was a lack of ethnic diversity in the participants interviewed. Understanding the potential barriers faced by underserved and underrepresented groups will be important for future implementation. We also note that the interviews and therapy were conducted during the COVID-19 pandemic. The context of restrictions, disruption, and anxiety may have affected the experience of agoraphobic avoidance and VR therapy. Finally, the lengthy interviews that covered multiple questions mean that it is likely that we may only have been able to capture a fraction of participants’ experiences in this report.

Acknowledgments
The authors thank the study participants and other people who worked on this study. They are grateful to the gameChange Lived Experience Advisory Panel for their valuable contribution to the study design, materials, and analysis. Members include Debbie Butler, Susie Booth, Len Demetriou, Zach Howarth, and Mary Mancini. gameChange was funded by the National Health Service National Institute for Health and Care Research (NIHR) Invention for Innovation program (Project II-C7-0117-20001). This work was also supported by the NIHR Oxford Health Biomedical Research Centre (BRC-1215-2000). DF is an NIHR Senior Investigator. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, or Department of Health. FW is supported by a Wellcome Trust Clinical Doctoral Fellowship (102176/B/13/Z).

Authors' Contributions
All the authors contributed to the study design. TK facilitated the LEAP. Data were collected by JB and AK. JB, AK, and LC conducted the primary analysis with contributions from DR and VP. FW and ML provided supervision. JB, FW, DF, and DR wrote the first draft of the manuscript. All the authors approved the final version of the manuscript.

Conflicts of Interest
DF is a founder and nonexecutive board director of Oxford VR, a University of Oxford spin-out company, which programmed and commercializes the gameChange treatment. DF holds equity in Oxford VR and receives personal payments. DF holds a contract for his university team to advise Oxford VR on treatment development. The University of Oxford, Oxford Health National Health Service Foundation Trust, and the McPin Foundation received a share of the licensing fee from Oxford VR for the gameChange software. SL does some consultancy work for Oxford VR. The other authors have no conflicts to declare.

Multimedia Appendix 1
Interview schedule for the gameChange experience study.
[DOCX File, 35 KB - games_v11i1e38065_app1.docx ]

Multimedia Appendix 2
Illustrative quotes by superordinate and subordinate theme.
[DOC File, 86 KB - games_v11i1e38065_app2.doc ]

References


Abbreviations

IPA: interpretative phenomenological analysis
LEAP: Lived Experience Advisory Panel
NHS: National Health Service
VR: virtual reality

©Jessica Bond, Alexandra Kenny, Vanessa Pinfold, Lisa Couperthwaite, The gameChange Lived Experience Advisory Panel, Thomas Kabir, Michael Larkin, Ariane Beckley, Laina Rosebrock, Sinéad Lambe, Daniel Freeman, Felicity Waite, Dan Robotham. Originally published in JMIR Serious Games (https://games.jmir.org), 16.01.2023. 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.
Engagement With Gamification Elements in a Smoking Cessation App and Short-term Smoking Abstinence: Quantitative Assessment

Nikita B Rajani¹,², PhD; Luz Bustamante³,⁴, MSci; Dominik Weth², MSc; Lucia Romo³,⁵, PhD; Nikolaos Mastellos¹, PhD; Filippos T Filippidis¹, PhD

¹Department of Primary Care and Public Health, Imperial College London, London, United Kingdom
²NextStage Consulting, Dubai, United Arab Emirates
³Laboratoire EA 4430-Clinique Psychanalyse Developpement, Department of Psychology, University of Paris Nanterre, Paris, France
⁴Research and Development Department, Kwit SAS, Strasbourg, France
⁵Inserm–Le Centre de Recherche en Epidémiologie et Santé des Populations 1018 UPS, Hôpital Raymond-Poincaré, Paris, France

Corresponding Author:
Nikita B Rajani, PhD
Department of Primary Care and Public Health
Imperial College London
St Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 7427615928
Email: nikita.rajani14@imperial.ac.uk

Abstract

Background: Gamification in smoking cessation apps has been found to improve cognitive outcomes associated with higher odds of quitting. Although some research has shown that gamification can also positively impact behavioral outcomes such as smoking cessation, studies have largely focused on physical activity and mental health. Only a few studies have explored the effects of gamification on smoking cessation outcomes, of which the majority have adopted qualitative methodologies and/or assessed engagement with apps using self-report.

Objective: This study aimed to explore levels of user engagement with gamification features in a smoking cessation app via in-app metrics. Specifically, the objective of this paper was to investigate whether higher engagement with gamification features is associated with the likelihood of quitting in the short term.

Methods: Data from a larger online study that recruited smokers seeking to quit were analyzed to address the objectives presented in this paper. The study took place between June 2019 and July 2020, and participants were primarily recruited via social media posts. Participants who met the eligibility criteria used 1 of 2 mobile apps for smoking cessation. In-app metrics shared by the developer of one of the smoking cessation apps, called Kwit, were used to assess engagement with gamification features. Out of 58 participants who used the Kwit app, 14 were excluded due to missing data or low engagement with the app (ie, not opening the app once a week). For the remaining 44 participants, mean (SD) values were calculated for engagement with the app using in-app metrics. A logistic regression model was used to investigate the association between engagement with gamification and 7-day smoking abstinence.

Results: In total, data from 44 participants who used the Kwit app were analyzed. The majority of participants were male, married, and employed. Almost 30% (n=13) of participants self-reported successful 7-day abstinence at the end of the study. On average, the Kwit app was opened almost 31 (SD 39) times during the 4-week study period, with the diary feature used the most often (mean 22.8, SD 49.3). Moreover, it was found that each additional level unlocked was associated with approximately 22% higher odds of achieving 7-day abstinence after controlling for other factors such as age and gender (odds ratio 1.22, 95% CI 1.01-1.47).

Conclusions: This study highlights the likely positive effects of certain gamification elements such as levels and achievements on short-term smoking abstinence. Although more robust research with a larger sample size is needed, this research highlights the important role that gamification features integrated into mobile apps can play in facilitating and supporting health behavior change.
KEYWORDS
gamification; smoking cessation; smoking abstinence; mHealth; mobile apps; mobile phone; smartphone; digital health; user engagement; cognitive outcome; self-support; in-app metrics

Introduction

Despite steps taken to tackle the global tobacco epidemic, smoking-related health disability and mortality remain concerning [1]. While the majority of smokers want to quit smoking, research shows that in many countries such as the United Kingdom, long-term abstinence rates remain low [2]. In addition to low smoking cessation rates, face-to-face access to cessation services has been falling in several countries [3]. Digital health solutions, such as mobile apps, have been found to be effective methods of reaching individuals unwilling or unable to access in-person services. However, low engagement and retention are common challenges for mobile apps. The integration of gamification, the use of game elements in a nongame context [4], has been found to be positively associated with higher app engagement [5,6]. Some examples of gamification elements, also known as on-screen features or tactics, include goal setting, levels and badges, progress tracking, and progress sharing. Since gamification shares key elements with behavior change theories, it is often applied to health behavior change interventions [6].

Prior research shows that engagement with gamification elements in smoking cessation apps can positively impact cognitive constructs vital for abstinence such as self-efficacy and motivation to quit [7]. Although research also shows the positive effects of gamification on behavioral outcomes, the majority of existing studies have focused on physical activity and mental health [8]. A few studies that explored the association between engagement with gamification and smoking cessation adopted qualitative methodologies and assessed engagement using self-reported data rather than objective app usage metrics [6,7]. We aimed to explore the level of user engagement with gamification elements in a smoking cessation app via the analysis of in-app metrics. We also aimed to investigate whether engagement with gamification elements is associated with the likelihood of achieving short-term smoking abstinence.

Methods

Study Overview

The data used for this paper's analyses were collected as part of a larger online study that took place from June 2019 to July 2020 and explored the effects of gamification on cognitive constructs vital for smoking cessation [7]. Participants were recruited via posters in public places in London, United Kingdom, and posts on various social media channels such as Facebook and Instagram. Interested participants were asked to fill out a screening questionnaire to assess eligibility. Eligible participants (Table S1, Multimedia Appendix 1) were assigned 1 of 2 smoking cessation apps; this paper focuses on users of Kwit, a gamified mobile app designed to help smokers quit smoking using cognitive behavior therapy [9]. Participants had to use the app at least once a week for 4 weeks and complete a questionnaire before (baseline), 2 weeks after (mid-study), and 4 weeks after app use (end of study). In-app metrics shared by the app developer were used to assess engagement with features. Out of the 70 participants who were assigned Kwit, 58 (83%) completed the study. A participant was considered to have completed the study if they self-reported to have engaged with the app at least once a week for the duration of the study and completed all questionnaires. Aside from free app access, participants had a chance to win a £50 (US $60) Amazon voucher.

The Kwit App

Kwit is a mobile app, developed by Kwit.SAS, that aims to help smokers quit and successfully remain abstinent from smoking. It uses cognitive behavior therapy principles and gamification to assist individuals [9]. The app consists of features such as a smoking diary to log cravings and triggers, motivation cards, and a calculator/tracker to monitor self-progress in relation to key achievements. Figure 1 presents screenshots of the app, showing the smoking diary feature, an example of a motivation card, the achievements tracking page, and how users can track cravings. App versions 4.1 to 4.4 (June 2019 to July 2020) were used by participants.
Measures
Sociodemographic measures included age (18-29 years, 30-41 years, 42-53 years, and 54-65 years), gender (male or female), marital status (single, married, or civil partnered), and employment status (unemployed: individuals willing or able to work but not employed; employed; nonemployed: students, individuals unable to work, and homemakers). Similar to other studies, the Fagerström questionnaire was used to assess nicotine dependence with responses categorized as low (0-4 points), moderate (5-7 points), and high (8-10 points) [10,11]. Additionally, to measure abstinence, participants were asked at the end of the study, “Have you smoked at all in the past seven days?” to assess the 7-day point prevalence of smoking abstinence. Participants who responded “No, not even a puff” were considered short-term quitters. Participants who selected other responses (“Yes, just a few puffs”; “Yes, between 1 and 5 cigarettes”; and “Yes, more than 5 cigarettes”) were categorized as smokers.

Aside from self-reported data, objective in-app metrics were also used in the analysis. Kwit SAS routinely collects and maintains a database of app usage statistics, including various user interactions with specific features and the app overall. Kwit SAS provided in-app metrics for the participant identification numbers shared with them. The metrics included the number of times the app was opened, levels completed, smoking diaries logged, achievements unlocked, and motivation cards viewed. Additional information on the in-app metrics is presented in Table S2 (Multimedia Appendix 1).

Statistical Analysis
The statistical software Stata 16 (StataCorp) was used for the analysis. Descriptive statistics were used to present general participant characteristics, nicotine dependence, and 7-day abstinence. Mean (SD) values were calculated for the in-app metrics. Unadjusted logistic regression models were run for each in-app metric to explore its association with 7-day abstinence. The specifications of the adjusted logistic regression model were based on an iterative process that considered collinearity. Significance was set at the 5% level ($P<.05$), and 95% CIs were presented.

Ethics Approval
The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly Declaration of Helsinki 1964 and later versions. Ethical approval was obtained from the Joint Research Imperial College London Research Ethics Committee prior to the beginning of the study (reference: 19IC5158).

Results
Among the 58 participants who completed the study (self-reported app usage once a week over the study duration and completion of all questionnaires), 14 participants were excluded from the analysis due to issues with the in-app metrics (eg, missing or inconsistent data; $n=9$) or due to inadequate engagement according to the in-app data (ie, the app was not used once a week; $n=5$). Table 1 shows that more than half of Kwit app users were 18 to 29 years of age ($n=24$, 55%). A majority were male ($n=27$, 61%), married ($n=35$, 80%), and employed ($n=28$, 64%). Moreover, the majority of participants had low to moderate dependence on nicotine ($n=42$, 95%). Almost a third ($n=13$, 30%) reported at the end of the study that they successfully abstained from smoking in the past 7 days.

As seen in Table 2, Kwit users opened the app almost 31 (SD 39) times on average over the 4-week study period. Among the metrics that were collected, the most frequently used features were logging of smoking diaries (mean 22.8, SD 49.3) and unlocking of achievements (mean 22.3, SD 16.5). Additionally, over the study period, motivation cards were opened on average 8.0 (SD 11.2) times and 7.7 (SD 4.9) levels were unlocked by Kwit users.
In the adjusted logistic regression model (Table 3), each additional level unlocked was associated with approximately 22% higher odds of achieving 7-day abstinence after controlling for other factors such as age and gender (odds ratio 1.22, 95% CI 1.01-1.47). The number of diaries logged and motivation cards opened were not significantly associated with 7-day abstinence in the adjusted model. The number of achievements was not included in the model since it was highly correlated with the number of levels. However, when the logistic regression model was rerun with the number of achievements instead of levels, the results suggested that each interaction with the achievements feature was associated with a 7% increased likelihood of reporting 7-day abstinence (95% CI 1.01-1.15).

Table 1. Characteristics of the study sample.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (N=44), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>18-29 years</td>
<td>24 (55)</td>
</tr>
<tr>
<td>30-41 years</td>
<td>13 (30)</td>
</tr>
<tr>
<td>42-53 years</td>
<td>4 (9)</td>
</tr>
<tr>
<td>54-65 years</td>
<td>3 (7)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27 (61)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (39)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>9 (21)</td>
</tr>
<tr>
<td>Married or civil partnered</td>
<td>35 (80)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>28 (64)</td>
</tr>
<tr>
<td>Nonemployed</td>
<td>13 (30)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Nicotine dependence</strong></td>
<td></td>
</tr>
<tr>
<td>Low (0-4 points)</td>
<td>25 (57)</td>
</tr>
<tr>
<td>Moderate (5-7 points)</td>
<td>17 (39)</td>
</tr>
<tr>
<td>High (8-10 points)</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>7-day smoking abstinence</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (30)</td>
</tr>
<tr>
<td>No</td>
<td>31 (70)</td>
</tr>
</tbody>
</table>

Table 2. Summary of engagement with mobile app features after 4 weeks of app use.

<table>
<thead>
<tr>
<th>In-app metric</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of times the app was opened</td>
<td>30.8 (39)</td>
</tr>
<tr>
<td>Number of motivation cards opened</td>
<td>8 (11.2)</td>
</tr>
<tr>
<td>Number of achievements unlocked</td>
<td>22.3 (16.5)</td>
</tr>
<tr>
<td>Number of diaries logged</td>
<td>22.8 (49.3)</td>
</tr>
<tr>
<td>Number of levels completed</td>
<td>7.7 (4.9)</td>
</tr>
</tbody>
</table>
### Table 3. Logistic regression investigating the association between gamification and 7-day smoking abstinence at the end of the study (N=44).

<table>
<thead>
<tr>
<th>Variable</th>
<th>7-day smoking abstinence, OR&lt;sup&gt;a,b&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.98 (0.91-1.05)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>Number of diaries logged</td>
<td>2.27 (0.45-11.52)</td>
</tr>
<tr>
<td>Number of motivation cards opened</td>
<td>0.97 (0.91-1.03)</td>
</tr>
<tr>
<td>Number of levels unlocked</td>
<td>1.03 (0.97-1.11)</td>
</tr>
<tr>
<td></td>
<td>1.22 (1.01-1.47)</td>
</tr>
</tbody>
</table>

<sup>a</sup>OR: odds ratio.
<sup>b</sup>Adjusted for all the variables included in the table.
<sup>c</sup>Ref: referent.

### Discussion

#### Principal Findings

We found that after 4 weeks of app use, almost 30% of smokers reported 7-day abstinence. This rate is generally within the range reported by other mobile app studies [12]. However, direct comparisons are difficult due to differences between interventions and varying methods of measuring cessation. Mobile app interventions such as Kwit can be associated with an increased likelihood of abstinence compared to no assistance or using willpower alone [13]. This study opens the possibility of using smoking cessation apps to aid individuals who are not accessing face-to-face services [14].

The analysis also found that engaging with levels was significantly associated with 7-day abstinence. According to Gnauk et al [15], levels are important as they can function as a goal-setting tool that marks progression and signals accomplishment. Similarly, the achievements or badges feature was also found to be associated with an increased likelihood of reporting abstinence. Achievements and levels are similar as they both provide regular feedback to users and remind them of their successes. This can lead to an increase in perceived competence, which facilitates health behavior change [16]. Generally, the positive effect of providing regular feedback in both remote and face-to-face interventions is well established [17]. Aside from the importance of features such as levels and achievements, it was also found that the likelihood of reporting abstinence was not statistically different with regards to age and gender after adjusting for engagement with app features. This could imply that any effects that Kwit may have on smokers might not vary by age or gender. However, it is important to note that the small sample size could impact the generalizability of the findings.

#### Limitations

The observational nature of the study does not allow for causal inference; future research could carry out a rigorously designed randomized controlled trial with apps that differ only with regards to the type and number of gamification features. Aside from the sample size, the attrition of participants from 70 to 44 participants could pose a threat to the internal validity of the study and limit the generalizability of the findings. Additionally, since research has shown that relapses in abstinence can occur over a longer period of time, the relationship between engagement with gamification features and long-term quitting cannot be determined; future research could consider having a longer follow-up period [18,19]. Furthermore, validating smoking abstinence using biochemical verification is the gold standard and would be recommended to increase the robustness of future research studies.

Although the use of in-app metrics in mobile app studies is sparse and provides an objective method of assessing engagement with app features, it may not have captured the full engagement experience of users. For example, we were able to assess the number of times users engaged with specific gamification elements but not the length of engagement. While this is an imperfect measure of engagement, it can be considered a more objective method (compared to self-report) that is not frequently adopted in mobile app studies.

#### Conclusions

The overall learnings from this research highlight that features such as levels and achievements can positively impact short-term smoking abstinence. While further investigation is warranted with a larger, diverse sample and a longer follow-up period, our findings have positive implications for the use of gamification in mobile apps to support behavioral outcomes such as smoking cessation.

### Acknowledgments

This research did not receive a grant or funding from any agencies that are public, commercial, or within the not-for-profit sector. Kwit provided free access to the study participants but had no other financial input.
Authors' Contributions
NBR, NM, and FTF were involved in the study design. NBR collected the necessary data and drafted the manuscript. DW and LB contributed to the data analysis and manuscript write-up. All authors read and approved the final manuscript.

Conflicts of Interest
LB undertakes consultancy and research for the Kwit app (Kwit SAS) in a PhD contract involving the National Association of Research and Technology and the University of Paris Nanterre. The other authors have no conflicts to declare.

Multimedia Appendix 1
Eligibility criteria and in-app metrics. [DOCX File, 14 KB - games_v11i1e39975_app1.docx]

References
7. Rajani NB, Mastellos N, Filippidis FT. Self-efficacy and motivation to quit of smokers seeking to quit: quantitative assessment of smoking cessation mobile apps. JMIR Mhealth Uhealth 2021 Apr 30;8(5):e17458 [FREE Full text] [doi: 10.2196/25030] [Medline: 33929336]


A Serious Game About Hematology for Health Care Workers (SUPER HEMO): Development and Validation Study

Julien Perrin\textsuperscript{1,2}, PhD, PharmD; Amélie Meeus\textsuperscript{1}; Julien Broseus\textsuperscript{2,3}, MD, PhD; Pierre-Jean Morieux; Valentine Di Ceglie; Julien Gravoulet\textsuperscript{1}, PharmD; Maud D'Aveni\textsuperscript{2,3}, MD, PhD

\textsuperscript{1}Faculté de Pharmacie, Université de Lorraine, Nancy, France
\textsuperscript{2}Centre Hospitalier Régional Universitaire, Nancy, France
\textsuperscript{3}Faculté de Médecine, Maïeutique et Métiers de la Santé, Université de Lorraine, Vandoeuvre-Lès-Nancy, France

Corresponding Author:
Maud D'Aveni, MD, PhD
Faculté de Médecine, Maïeutique et Métiers de la Santé
Université de Lorraine
9 avenue de la Forêt de Haye
Vandoeuvre-Lès-Nancy, 54505
France
Phone: 33 383153289
Email: m.daveni-piney@chru-nancy.fr

Abstract

Background: Complete blood count (CBC) and hemostatic screening tests are among the most commonly prescribed blood tests worldwide. All health care workers (nurse practitioners, pharmacists, dentists, midwives, and physicians) are expected to correctly interpret the results in their daily practice. Currently, the undergraduate hematology curriculum consists predominantly of lecture-based teaching. Because hematology combines basic science (blood cells and hemostasis physiology) and clinical skills, students report that they do not easily master hematology with only lecture-based teaching. Having interviewed students at the University of Lorraine, we considered it necessary to develop new teaching approaches and methods.

Objective: We aimed to develop and validate a serious game about CBC analysis for health care students. Our primary objective was to help students perceive hematology as being a playful and easy topic and for them to feel truly involved in taking care of their patients by analyzing blood tests. We considered that this game-based approach would be attractive to students as an addition to the classic lecture-based approach and improve their knowledge and skills in hematology.

Methods: We developed an adventure game called SUPER HEMO, a video game in which the player assumes the role of a protagonist in an interactive story driven by exploration and problem-solving tests. Following validation with beta testing by a panel of volunteer students, we used a novel, integrated teaching approach. We added 1.5 hours of gaming to the standard curriculum for a small group of volunteer students. Physician and pharmacy students in their third year at a single French university were invited to attend this extracurricular course. Pregame and postgame tests and satisfaction surveys were immediately recorded. Final hematology exam results were analyzed.

Results: A total of 86 of 324 physician students (26.5\%) and 67 of 115 pharmacy students (58\%) opted to participate. Median scores on the pre- and posttests were 6 out of 10 versus 7 out of 10, respectively, for the physician students, (\textit{P}<.001) and 7.5 out of 10 versus 8 out of 10, respectively, for the pharmacy students (\textit{P}<.001). At the final hematology evaluation, physician students who played SUPER HEMO had a slightly better median score than those who did not: 13 out of 20 versus 12 out of 20, respectively (\textit{P}=.002). Pharmacy students who played SUPER HEMO had a median score of 21.75 out of 30; this was not significantly different from pharmacy students who did not play SUPER HEMO (20/30; \textit{P}=.12). Among the participants who answered the survey (n=143), more than 86\% (123/143) believed they had strengthened their knowledge and nearly 80\% (114/143) of them had fun.

Conclusions: Feedback from this game session provided evidence to support the integration of interactive teaching methods in undergraduate hematology teaching. The development of SUPER HEMO is intended to be completed so that it can become a support tool for continuing education.

\textit{(JMIR Serious Games 2023;11:e40350)} doi:10.2196/40350

https://games.jmir.org/2023/1/e40350
### Introduction

Numerous serious games (SGs) have been developed to improve nursing [1] and medical [2] knowledge and skills. The main subjects are surgery [3-5], emergency medicine [6,7], pharmacy [8] for health care students, and other subjects, such as preventive medicine for adolescents and young people [9-12]. Systematic reviews of SGs conclude that they seem to be at least as effective as other digital education modalities [13], but pedagogical effectiveness, participant behavior, and patient health outcomes have barely been evaluated [14-16].

In the context of the global COVID-19 pandemic, the demand for online learning increased worldwide. However, it has been reported that reduced peer and teacher interaction can cause motivation issues [17,18]. Recently, online learning has been described negatively, especially by health care students in clinical practice [19]. Many health care students see working in the health care setting as a vocation rather than a job, with patient-centered and compassionate care being the basis of this view [20]. Therefore, video games that enable specifically situated, experiential learning by introducing different unwell characters represent an attractive learning tool for today’s students, who are very receptive to computer-based learning.

SGs for hematology education are rare and deal only with transfusion [21,22]. One important part of hematology education is complete blood count (CBC). CBC is a quantitative and qualitative evaluation of blood cells and stands as one of the most common laboratory tests in medicine, being indicated for a vast number of conditions. CBC interpretation is taught only with lecture-based courses in hematology. An internal survey among physician and pharmacy students at the University of Lorraine revealed that more than 50% of students considered CBC interpretation hard to master. Because correctly applying CBC knowledge to form patient diagnoses and make clinical decisions is a requirement for all health professionals, we decided to implement an SG for hematology education.

We developed an SG (Figure 1) named SUPER HEMO with a structured, 3-phase development framework that included preparation and design, development, and formative evaluation, as described previously [23].

![SUPER HEMO development](https://games.jmir.org/2023/1/e40350)

### Methods

#### Preparation and Design

**Framework**

As a reference for the creation and development of the game, we took inspiration from role-playing and adventure games (eg, point-and-click games and visual novels), as these are commonly played at home by students. The game’s contents, including a database with questions, answers, and feedback, were based on knowledge from the Collège national des enseignants en hématologie (the French national college of hematology teachers).

At first, we created an educational committee composed of 3 hematology experts (1 pharmacist and 2 medical doctors), a pharmacist with gamification expertise (with no expertise in hematology), and an instructional designer. Meetings took place every month for the first 6 months of the project. The goals of this committee were to define pedagogical objectives, game design, and game modes, and to write the clinical cases’ dialogue. During the writing of the clinical cases, physician students volunteered to test some of the cases before their integration into the game. The committee also looked for funding. The raised funds were mainly used to hire a graphics designer and a game developer.

Then, a working committee was created with 2 medical doctors, a pharmacist with hematology expertise, a pharmacist with gamification expertise, an instructional designer, a graphics designer, and a game developer. Meetings occurred every 2 months to discuss practical issues arising from game conception.
Some members worked together outside of the meetings in groups, such as the graphics designer with the hematology expert or the game developer with the instructional designer and the hematology expert.

The project was then presented to the pedagogical council of each faculty.

**Story**

SUPER HEMO is set in a dreamlike “Red Cell World” (Multimedia Appendix 1), with red-blood-cell trees and depictions of the organs involved in erythropoiesis, including a “lung mountain,” “medullar cave,” “kidney rock,” “spleen fortress,” and “thyroid isthmus.”

Players can choose a female or a male avatar on the home screen. They then complete the introduction, in which “Lady Stem Cell” explains the world and the game’s settings and instructions and gives the player the “CBC asset,” which is the power to check the blood parameters of the unwell characters encountered and interpret the corresponding results. Players assume the role of a hematology superhero named SUPER HEMO. SUPER HEMO can meet 5 unwell characters in 5 different steps. The player must answer their questions and find the best way to diagnose and cure them. “Magicians” (radiologists, pathologists, hematology-biologists, pharmacists, and geneticists) can be summoned to help the player find the right answer (Multimedia Appendix 2) in exchange for gold coins.

**Mechanics**

To complete the world’s challenges (Figure 2), SUPER HEMO must explore 5 clinical cases.

![Figure 2. “Red World” mechanics.](image)

Each stop (represented by a light blue button) corresponds to the clinical case of an unwell character. Players choose to explore the clinical cases in the order that they decide. When SUPER HEMO successfully ends a case, players are rewarded with 1, 2, or 3 stars depending on the number of errors they made. Moreover, a hidden hematopoietic cell can be caught to unlock a minigame to earn gold coins. A wrong answer to a skill question makes SUPER HEMO lose the patient’s trust and lose gold coins, and a wrong answer to a treatment question makes the hero lose the patient’s trust. Lady Stem Cell gives immediate feedback to every question, thus maintaining the user’s motivation and commitment to the game [24]. If all gold coins and the patient’s trust are lost, Mister Insurance takes SUPER HEMO back to the beginning of the case, where the player can immediately play again if they have enough gold coins; otherwise, they can still earn gold coins in minigames before trying the case again. The hidden collected cells constitute an atlas of hematopoietic cells that contains a precise description and a real picture (taken by microscopy) of each blood cell. Minigames are always related to hematology in a fun way.

When SUPER HEMO reaches 80% success in the world’s 5 clinical cases (and achieves a total of 12 stars of a possible 15) the player wins a new asset (a myelogram) and is thereby allowed to move on to a higher level, which corresponds to another hematopoietic world.

Throughout the game, to maintain the dreamlike atmosphere and immersion, different theme songs accompany the player (with the possibility to mute them). The player also has the option to reset the game on the home screen (not pictured).

**Development**

**Technology**

The game was programmed in January 2019 using C# in the Unity game engine, for computers only (ie, there was no mobile app). The invention was protected with an Inter Deposit Digital Number on September 9, 2020, for the University of Lorraine.
To date, the game is available in French only and is free of charge at the University.

**Video Game Beta Tests**

Three beta-test sessions were organized with small groups of undergraduate and graduate physician and pharmacy volunteer students. The goal was to obtain a primary evaluation of the game design and concept and to debug the game. The volunteers played the game for 1 hour and were then asked to fill in a questionnaire (Multimedia Appendix 3) about several aspects of the game, including the graphical interface, gameplay, use of multimedia, and educational content, and provide ideas for improvement and general comments. We collected the questionnaires and analyzed the answers to modify or improve the game and the teaching methods, if required.

**Formative Evaluation**

**Ethics Approval**

The study was conducted in accordance with the Helsinki Declaration and Resolution. The study was approved by the Pedagogic Committee of the Faculté de Pharmacie, Université de Lorraine (June 17, 2021) and Faculté de Médecine, Université de Lorraine (July 7, 2021).

**Recruitment and Game Evaluation**

Students in their third year of medicine and pharmacy courses were identified as the most appropriate study participants, since hematology is a regular and mandatory course unit during this academic year. These student cohorts were expected to benefit the most from additional exposure to clinical learning while using SUPER HEMO, as their upcoming final exams were planned to take place shortly after game exposure. The game evaluation was performed after 2 weeks of classical hematology teaching. The gaming course consisted of 1.5 hours of gaming (in the Red World) that covered the following standard red blood cell disorders: anemia and polycythemia. The game evaluation was split into 2 phases examining (1) the immediate knowledge acquisition of the players and (2) whether the players successfully completed their final hematology examination.

First, to evaluate the effectiveness of this teaching method, voluntary participants from each course were asked to complete a knowledge test consisting of 10 multiple-choice questions (Multimedia Appendix 4); one point was obtained for each correct answer. The participants completed the 10-question test before (pretest) and after playing the game for 1 hour (posttest). An online questionnaire was designed to assess playability and the students’ understanding of SUPER HEMO. Students were also asked to rate their level of confidence after the gaming session on a questionnaire. Qualitative data considering the students’ general feedback was also collected. All answers were anonymized.

Second, successful completion of the final hematology examination was extracted for students who had participated in SUPER HEMO and those who had not by the pedagogic committee, and mean results were compared.

**Statistics**

Statistical analyses were carried out using Prism (version 5.0; GraphPad). Comparisons of participation rates for the physician and pharmacy students, as well as the female to male ratio of the groups, were made with the Fisher exact test. Pre- and posttest scores and the final evaluation were expressed as median values, with the range and 25th to 75th percentiles. Pre- and posttest results were compared using a paired Wilcoxon signed-rank test in the 3 groups (ie, the overall population, the physician students, and the pharmacy students). Final evaluation results were compared using the Mann-Whitney test. As participation was on a voluntary basis, a post hoc power analysis of the final evaluation’s scores was also performed.

**Results**

**Beta Tests**

The results of the 3 beta tests (Figure 3) indicated a significant interest in this new SG.

The interface received a score of 3.7 of 5. As for the multimedia aspect, the graphics and music received scores of 4.1 and 3.7 of 5, respectively. All students stated that they enjoyed the game, 86% (19/22) found the game fun to use, and 68% (15/22) even lost track of time while playing. Regarding the educational content, 90% of the students (20/22) found that SUPER HEMO represented an efficient method to learn hematology.

The beta tests allowed us to detect a few points to improve SUPER HEMO and address issues encountered by the students. Some students had difficulties answering certain types of questions (for example, ones that used drag-and-drop), or they could not find the minigames. This led us to integrate tutorials at the beginning of each game, such as on how to answer questions encountered for the first time. We also modified some questions and minigame rules to better drive the players. Moreover, some players forgot the CBC results while talking to the unwell characters and frequently wanted to check them; therefore, we added a button to allow the players to check the CBC results as often as they wanted to. Finally, as the game tasks could be interrupted, we added a “log info” button to remind the students of dialogue text.

The free comments included mostly thanks for the initiative and anticipation for the sequels.
Integrated Teaching Approach in Addition to the Standard Hematology Undergraduate Curriculum

A total of 153 volunteer students were recruited, including 86 of 324 physician students (26.5%) and 67 of 115 pharmacy students (58%), who agreed to participate in this complementary session at the end of the standard lecture-based hematology course. Of note, the proportion of female students was higher in the participant group (Table 1).

Overall, the volunteer students were evaluated on a 10-point scale (Table 2 and Figure 4). They had a higher posttest score (median 7.5, range 2.5-10) than pretest score (median 6.5, range 1.5-9; *P* = .001), indicating that the game slightly improved their immediate knowledge acquisition. The median pre-and posttest scores were 6 (range 1.5-8) and 7 (range 2.5-10), respectively, for the physician students (*P*<.001) and 7.5 (range 3-9) and 8 (range 3.5-10), respectively, for the pharmacy students (*P*<.001).

At the final examination (Figure 4), we observed that the physician students who played SUPER HEMO obtained a slightly higher score (median 13/20, range 6-17) than those who did not play SUPER HEMO (median 12/20, range 5-17; *P*=.002), with a satisfactory study power (83%). The pharmacy students who played SUPER HEMO had a score (median 21.75/30, range 12.25-27.25) that was not statistically significantly different from those who did not play SUPER HEMO (median 19.8/30, range 9.5-26.75; *P*=.12). Unfortunately, the study power was low (64%) for this group.

Table 1. Comparison of participants and nonparticipants in SUPER HEMO evaluation.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants, n</th>
<th>Nonparticipants, n</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All students (N=439)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>110</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Physician students (n=324)</td>
<td></td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>Female</td>
<td>60</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>26</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>Pharmacy students (n=115)</td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Female</td>
<td>50</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Pre- and posttest evaluations (scores are on a 10-point scale).

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median score</td>
<td>Range</td>
<td>25th-75th percentile</td>
</tr>
<tr>
<td>All students</td>
<td>6.5</td>
<td>1.5-9</td>
<td>5.5-8</td>
</tr>
<tr>
<td>Physician students</td>
<td>6</td>
<td>1.5-8</td>
<td>5-7</td>
</tr>
<tr>
<td>Pharmacy students</td>
<td>7.5</td>
<td>3-9</td>
<td>6.5-8.5</td>
</tr>
</tbody>
</table>
Figure 4. SUPER HEMO evaluation.

**Pre and post-test evaluation**

![Box plots showing pre and post-test scores](image)

**Final examination**

![Box plots showing final examination scores](image)

**Gameplay Satisfaction**

Among 153 volunteer students, 143 answered the questionnaire (93.4%). Game experience received a score of 4.7 out of 5, and 79% of the students (n=113) found the game fun to use. Regarding multimedia, graphics received a score of 4.7 out of 5, and music received 3.9 out of 5. The minigames were scored as attractive (4.1/5; Figure 5).

Concerning the educational content, most of the 153 students indicated that they had augmented their knowledge (n=130, 91%), had made progress in hematology (n=137, 96%), and better understood their courses (n=124, 87%) after playing SUPER HEMO. In addition, 71 students (50%) indicated that they specifically aimed to obtain the top 3 stars at the end of each case in the game, and 70 students (49%) dedicated specific effort to collect all the hidden cells. While the effect of rewards on memory appears well documented, it has recently been reported that incentives can also have counterproductive effects on memory [25]. Our videogame reward system was, however, developed to create a realistic game environment with well-known reinforcement and reward schedules. Of the 153 participants, 106 (74%) indicated that the flow of the game suited their knowledge, zero indicated they were bored by the game, 5 (3%) became lost in the game, and only 1 gave up.
Discussion

Principal Findings

SUPER HEMO was developed to increase students’ motivation for learning hematology. In fact, as previously described, when students feel involved, they are more likely to achieve educational goals [26]. But with only 4 hematology teachers at the university, bedside (for clinical symptoms) and laboratory (for cell recognition) teaching in undergraduate medical education could not be fully accomplished for that year’s 324 physician and 115 pharmacy students. Consequently, SUPER HEMO was developed to confront students with different clinical situations, improve their cell observation skills, and supplement the classical teaching model. Because players may have different skills, we developed the pedagogic content concomitantly with gamification [27]. Special characters, such as Lady Stem Cell, were created to regularly debrief the players on wrong or right answers to questions [28]. SUPER HEMO level 1 (the Red World) was developed and beta tested over a 2-year period. Subsequently, SUPER HEMO’s approach was integrated to the standard hematology undergraduate curriculum of third-year physician and pharmacy students and the SG was evaluated.

In health care education, studies comparing SGs to other teaching methods with prospective evaluations are scarce [29,30]. We assumed that this prospective evaluation might be useful to justify SUPER HEMO’s integration with the standard hematology curriculum. The satisfaction questionnaire clearly showed that most students enjoyed playing SUPER HEMO (113/142, 80%) and felt that they learned from it (123/142, 86%). These results demonstrate the students’ strong commitment to the game. This result was reinforced by the free comments (eg, “I find the interface truly attractive; it really makes me want to do well” and “A good experience, dialogs or context are quite funny but still remain consistent!”). Several students inquired about playing the rest of the game (eg, “We’re looking forward to the white and yellow worlds’ release” and “I hope I can play again soon!”). Students highlighted the link between the game and seriousness (eg, “It allowed me to study in a fun way” and “Fun way to learn hematology or to practice without the feeling that I have studied”).

Knowledge improvement with this complementary method to the standard course was more difficult to evaluate. If a positive effect of SUPER HEMO was clearly observed for short-term knowledge (ie, the results of the pre- and posttests), the measurement of learning outcome (ie, the results of the final examination) should also be discussed. First, the first SUPER HEMO session ran through the COVID-19 pandemic, when restrictions were in place on attending courses. Therefore, we could not evaluate the effectiveness of this SG with a randomized controlled trial. After discussion with the pedagogic committee during the pandemic, courses were transmitted virtually to students who wanted to stay at home; less than 10% of students attended on site. In order not to penalize medically or psychologically fragile students, we decided to invite students...
to participate as they chose to do so. We can hypothesize that volunteer students who came to the SUPER HEMO session were the most motivated in the class, thus biasing the results of the final exams. Second, we observed that pharmacy students were more likely to attend the SUPER HEMO session than the physician students. Of note, the pharmacy students were probably more motivated, as their final exam took place a few days after the gaming session, whereas the physician students had their final exam 1 month after the gaming session. For the pharmacy students, although the participation rate was satisfactory, the study was underpowered, suggesting that follow-up studies (within the next few years) with more students might confirm that playing SUPER HEMO allowed students to obtain a better score in the final hematology evaluation (as was observed for the physician students). Third, our conclusions on interest in SUPER HEMO should be moderated by consideration of the SUPER HEMO session design. The session had a time limit of 90 minutes and was based only on the available “Red World,” which dealt with anemia and polycythemia, while the final evaluation contained questions on red blood cells, white blood cells, and platelets. We therefore propose that in the future (1) three worlds corresponding to the three lineages of hematopoietic cells in the CBC be made available to encompass the whole hematology program, (2) game-based learning should have no time restrictions, and (3) new evaluations with more participants should be organized with other universities.

Currently, we consider that our research contributes to the literature by providing a new game for teaching hematology and an investigation of the effectiveness of the SG context for hematology learning. This first experiment with SUPER HEMO commits us to develop this SG for hematology in future academic years. Future work will focus on developing levels 2, 3, and 4, the “White World,” “Yellow World,” and “Complex World,” respectively, with all parameters varying in the latter. We propose an evolving game that covers the entire program of hematology learning. This first experiment with SUPER HEMO was demonstrated beyond the initial beta-testing pilot study; we obtained preliminary evidence that SUPER HEMO might be a useful educational tool. We will use it as a supplement to lecture-based courses and trace the time and frequency of logins to SUPER HEMO. Last, we propose to correlate this continuous follow-up studies (within the next few years) with more students to participate as they chose to do so. We can hypothesize that volunteer students who came to the SUPER HEMO session were the most motivated in the class, thus biasing the results of the final exams. Second, we observed that pharmacy students were more likely to attend the SUPER HEMO session than the physician students. Of note, the pharmacy students were probably more motivated, as their final exam took place a few days after the gaming session, whereas the physician students had their final exam 1 month after the gaming session. For the pharmacy students, although the participation rate was satisfactory, the study was underpowered, suggesting that follow-up studies (within the next few years) with more students might confirm that playing SUPER HEMO allowed students to obtain a better score in the final hematology evaluation (as was observed for the physician students). Third, our conclusions on interest in SUPER HEMO should be moderated by consideration of the SUPER HEMO session design. The session had a time limit of 90 minutes and was based only on the available “Red World,” which dealt with anemia and polycythemia, while the final evaluation contained questions on red blood cells, white blood cells, and platelets. We therefore propose that in the future (1) three worlds corresponding to the three lineages of hematopoietic cells in the CBC be made available to encompass the whole hematology program, (2) game-based learning should have no time restrictions, and (3) new evaluations with more participants should be organized with other universities.

Currently, we consider that our research contributes to the literature by providing a new game for teaching hematology and an investigation of the effectiveness of the SG context for hematology learning. This first experiment with SUPER HEMO commits us to develop this SG for hematology in future academic years. Future work will focus on developing levels 2, 3, and 4, the “White World,” “Yellow World,” and “Complex World,” respectively, with all parameters varying in the latter. We propose an evolving game that covers the entire program of hematology learning. This first experiment with SUPER HEMO was demonstrated beyond the initial beta-testing pilot study; we obtained preliminary evidence that SUPER HEMO might be a useful educational tool. We will use it as a supplement to lecture-based courses and trace the time and frequency of logins to SUPER HEMO. Last, we propose to correlate this continuous training to student results for the third-year final exams, final graduation exams (at the end of the fifth year), and for the student’s specialty choice.

Acknowledgments
We deeply thank all participants and the University of Lorraine for supporting this study, especially Catherine Claus-Demangeon, Edwige Morin, and Hugo Bousquet. We also deeply thank Prof Beatrice Faivre, Prof Raphaël Duval, and Prof Marc Braun. This study was financed by Initiatives d’excellence en formation innovantes – Réseau des écoles de management et d’ingénierie de la santé (IDEFI-REMIS) (ANR #11-IDFI-0033) and the pharmacy and medicine faculties in Nancy.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Red Cell World.
[PPTX File, 1365 KB - games_v11i1e40350_app1.pptx]

Multimedia Appendix 2
SUPER HEMO avatars speaking with different characters.
[PPTX File, 1345 KB - games_v11i1e40350_app2.pptx]
References


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.
A Software Engineering Framework for Reusable Design of Personalized Serious Games for Health: Development Study

Stéphanie Carlier¹,², MSc; Vince Naessens¹, MSc; Femke De Backere¹,², PhD; Filip De Turck¹,², PhD

¹Internet Technology and Data Science Lab, Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium
²Interuniversity Microelectronics Centre, Ghent, Belgium

Corresponding Author:
Stéphanie Carlier, MSc
Internet Technology and Data Science Lab
Faculty of Engineering and Architecture
Ghent University
Technologiepark-Zwijnaarde 126
Ghent, 9052
Belgium
Phone: 32 9 331 49 38
Email: stephanie.carlier@ugent.be

Abstract

Background: The use of serious games in health care is on the rise, as these games motivate treatment adherence, reduce treatment costs, and educate patients and families. However, current serious games fail to offer personalized interventions, ignoring the need to abandon the one-size-fits-all approach. Moreover, these games, with a primary objective other than pure entertainment, are costly and complex to develop and require the constant involvement of a multidisciplinary team. No standardized approach exists on how serious games can be personalized, as existing literature focuses on specific use cases and scenarios. The serious game development domain fails to consider any transfer of domain knowledge, which means this labor-intensive process must be repeated for each serious game.

Objective: We proposed a software engineering framework that aims to streamline the multidisciplinary design process of personalized serious games in health care and facilitates the reuse of domain knowledge and personalization algorithms. By focusing on the transfer of knowledge to new serious games by reusing components and personalization algorithms, the comparison and evaluation of different personalization strategies can be simplified and expedited. In doing so, the first steps are taken in advancing the state of the art of knowledge regarding personalized serious games in health care.

Methods: The proposed framework aimed to answer 3 questions that need to be asked when designing personalized serious games: Why is the game personalized? What parameters can be used for personalization? and How is the personalization achieved? The 3 involved stakeholders, namely, the domain expert, the (game) developer, and the software engineer, were each assigned a question and then assigned responsibilities regarding the design of the personalized serious game. The (game) developer was responsible for all the game-related components; the domain expert was in charge of the modeling of the domain knowledge using simple or complex concepts (eg, ontologies); and the software engineer managed the personalization algorithms or models integrated into the system. The framework acted as an intermediate step between game conceptualization and implementation; it was illustrated by developing and evaluating a proof of concept.

Results: The proof of concept, a serious game for shoulder rehabilitation, was evaluated using simulations of heart rate and game scores to assess how personalization was achieved and whether the framework responded as expected. The simulations indicated the value of both real-time and offline personalization. The proof of concept illustrated how the interaction between different components worked and how the framework was used to simplify the design process.

Conclusions: The proposed framework for personalized serious games in health care identifies the responsibilities of the involved stakeholders in the design process, using 3 key questions for personalization. The framework focuses on the transferability of knowledge and reusability of personalization algorithms to simplify the design process of personalized serious games.

(JMIR Serious Games 2023;11:e40054) doi:10.2196/40054

https://games.jmir.org/2023/1/e40054

https://games.jmir.org/2023/1/e40054

JMIR Serious Games 2023 | vol. 11 | e40054 | p.35 (page number not for citation purposes)
KEYWORDS
serious game; health care; personalization; domain knowledge; framework; eHealth; cocreation

Introduction

Background

Serious games have a primary objective other than pure entertainment as they seek to educate or train, among others [1,2]. In many domains of health care, serious games have shown positive effects and are increasingly used to motivate treatment adherence, reduce treatment costs, and educate patients and their families on a specific pathology [3-7]. For example, serious games are used for the treatment of mental health disorders, such as anxiety disorders or depression [5,8-10], physical [11-14] or cognitive rehabilitation [15-17], or the education of health professionals and patients [18-22]. Multiple reviews exist on this emerging field that state that serious games can be an effective tool for health, but the research remains in its infancy, limited by design and evaluation challenges [5,10,20,21,23-25].

One of these challenges is the long-term and continuous support of the targeted users. The user’s abilities will not only evolve but also based on the user’s current context, a different configuration or approach might be called for. Many researchers have indicated the need for a personalized approach in serious games, abandoning the one-size-fits-all approach [26-31]. Serious games often fail to sustain long-term retention and treatment adherence if gamification mechanisms are not adaptive and cannot dynamically reengage the user [32,33]. Users lose motivation and games become predictable. Although the primary objective of a serious game is not entertainment, it remains crucial that the game is entertaining to retain user engagement [34]. To create long-term engagement in serious games, a balance between challenge and skill, leading to a state of flow, must be achieved [35].

Games tailored to the user generally result in better performance outcomes for the user, making personalization a key aspect of a successful serious game [36-39]. Different terms, such as adaptability, adaptivity, personalization, contextualization, and customization, are used in the literature to indicate the tailoring of serious games to users. Sajjadi et al [26] introduced the overarching term individualization when no distinction between these concepts is necessary. Adapting a serious game to the needs of a user can be done at design time, before starting the game (ie, static personalization), or while playing the game (ie, dynamic personalization).

Adaptability is defined by multiple researchers as the possibility to change an environment based on the user’s changing needs, whereas adaptivity is the dynamic or automatic adjustment of game elements to the individual’s actions or performance [26,29]. Personalization is characterized by the, often automatic, adaptation of the game based on the profile or context information of a specific individual user, such as heart rate or age [26]. Customization can be seen as changing the system based on the needs of a user group or an individual user, manually or automatically, and is often related to changes in appearance and content [26,40]. Streicher and Smeedinck [29] considered personalization as a specific form of customization, while they saw adaptability and adaptivity as a means to achieve personalization or customization. These concepts refer to the tailoring of the game at run time (ie, before or during gameplay), which differs from the player-centered design in which decisions are made during the design of the game based on the needs of a specific target group [41].

Not only do people learn in different ways and paces and perceive the difficulty level of the game differently, but the game itself is also experienced differently by different people, and not all game elements will work for everyone. In addition, the specific skills of the user might vary and develop over the course of playing the game [26,35,42,43]. The game should therefore be able to respond by adapting to the user, that is, ensuring a state of flow [4,35].

Flow theory models the relationship between the level of challenge in the game and the skill level of the user [44,45]. According to the Flow Model, the user is in a state of flow (ie, total immersion with maximized focus and performance) when the game has a clear goal and the user receives direct feedback on their performance related to this goal [45]. More importantly, to enter this state of flow, or the flow channel, the goal and related challenge level of the game should match the skill level of the user [4,29]. Frustration occurs when the game is too difficult for the perceived skill level and boredom sets in if the user is not challenged sufficiently by the game. Therefore, a serious game should maintain a balance between these parameters, even as the skill level of the users increases throughout the course of the game, to ensure that a state of flow is achieved, as shown in Figure 1. Because people differ and learn at different rates, the serious game should be able to detect and respond to the changing context and skill level of the user [4].

Another disadvantage of current gamified health applications is that the level of customization is often lacking, resulting in gamification that does not take health purposes, changes, or target groups into account [32]. Entertainment games already include different preferences among users by identifying player types using models such as the Bartle Model [46]. However, these models cannot be generalized to serious games because they are too limited. More context or player aspects need to be considered for the personalization of serious games, such as anxiety, stress, learning style, engagement, performance, skill level, and so on [26]. In contrast to the audience of entertainment games, the target audience of serious games is larger, including nongamer types.

Serious games and gamification, with a user-centered, adaptive, and personalized approach, show promise in increasing treatment adherence and boosting engagement with interventions [3]. Creating a personalized serious game from scratch for each type of user is a costly and challenging operation for developers and domain experts [17,29]. To create a serious game, expert knowledge of relevant domains is necessary. Often, therapists and domain experts are continuously involved in this design.
Some examples are as follows: the involvement of a team of therapists in the design process of a serious game for anxiety reduction in children with autism spectrum disorder [8]; conducting in-depth interviews with occupational therapists for a serious game for cognitive impairment [15]; and validating the content of an informational serious game on COVID-19 by consulting a specialized team of physicians, professors, and medical students [48].

Different methods exist to include personalization in serious games. Some approaches focus on classifying the user according to a certain player type [41,49,50], whereas others focus on changing game aspects [33,35,51]. Rule-based adaptation defines rules that, when satisfied, lead to predefined actions that determine the further course of the game. However, this can lead to a less effective game because adaptation options are limited by the predefined rules [4]. Plan-based adaptation can be considered as a collection of state machines. Each plan is a state machine of which each state is an executable action in the game. If a condition is met, the active states can be selected to determine the course of the game. This adaptation method can, for example, be used in games containing several storylines, unlocking different aspects of the story depending on the user’s current physiological state (eg, relaxed vs stressed states) [4]. Model-based adaptation allows the creation of models for various game elements. These models can dynamically change based on the changing user information. This method allows for more complex analysis and adaptation techniques such as the use of artificial intelligence to predict the progression of the user in the future [4,52]. Studies exist on how gamification and serious games can be personalized and which factors influence these decisions [53]. However, it remains unclear how this information can be integrated into the design and implementation of a fully personalized and adaptive serious game.

Research exists on frameworks for the design of serious games; however, these are often designed for specific health care domains, such as physical [13,54-56] or cognitive rehabilitation [15,57,58], or focus on specific technologies, such as emotion recognition and gamification patterns [59,60]; only a few focus on the development of serious games in general [28,61]. Nevertheless, these frameworks often remain vague on how personalization can be achieved, use one specific model for personalization, or do not target personalized serious games [13,60,61]. Moreover, only one study reported the need for reusable serious games components and proposed a serious game framework for reusable intelligent software components (eg, emotion recognition and learning algorithms) [62]. However, this study is limited to the implementation of software components and does not state how domain knowledge and intelligent personalization algorithms can be integrated into the multidisciplinary design process of serious games.

Owing to the lack of standardized frameworks for the design of a (personalized) serious game, the challenging nature of the design process, and the focus on using case-specific guidelines, the evaluation of serious games is often limited and varies widely as different approaches are taken. Research should aim to evaluate the effectiveness of serious games more rapidly and in a controlled setting, focusing on comparing approaches using the same environment [10,17].

**Objective**

Overall, when creating a personalized serious game, 3 questions need to be asked:

- Why is the game personalized?
- What parameters can be used for personalization?
- How is the personalization achieved?

In this study, we aimed to answer these questions by proposing a software engineering framework that streamlines the design process of personalized serious games and facilitates the reuse of domain knowledge and personalization algorithms to reduce development costs.

Currently, the development of personalized serious games is a costly and complex process that requires the continuous involvement of different stakeholders, such as (game)
developers, domain experts, and software engineers. Furthermore, this process must be completely repeated for each serious game and is often use-case specific. This means that each stakeholder should share their expertise throughout the design process of each new serious game, irrespective of the subject of the game, as their knowledge is never formalized or transferred to new serious games. This results in a complex process with many dependencies and redundancies between different stakeholders and serious games, which can be visualized as shown in Figure 2A. This approach complicates the evaluation of serious games and personalization strategies. With that many different parameters, such as design approach, target audience, or personalization algorithms, the comparison and evaluation of different personalization strategies are limited. Moreover, very little is known about how personalization strategies can be applied to serious games, as research is focused on the development of serious games for specific use cases, without attention to reusability.

This study aims to reduce these disadvantages by proposing a framework to simplify this process. The framework streamlines the development process of personalized serious games by considering the value of cocreation with different stakeholders, as shown in Figure 2B. However, by focusing on the reusability and formalization and transferability of expert knowledge, the dependencies between the stakeholders can be decoupled, thereby reducing the development cost of a personalized serious game.

The following paragraphs are structured as follows. The Methods section describes the Use and Design of the Software engineering framework, followed by the Generic Framework, which discusses the software engineering framework and the responsibilities of the involved stakeholders. To assess the proposed framework, a simple proof of concept was implemented, that is, an existing game was first transformed into a serious game, which is explained in the Methods section in A Serious Game for ShoulderRehabilitation. Next, the possibilities of the framework were illustrated by transforming this framework into a personalized serious game in the section Proof of Concept: an Adaptive and Personalized SeriousGame, followed by a discussion of the evaluation of the resulting serious game in the Results section. Finally, the conclusions of this study are discussed in the Discussion section.

Figure 2. (A) A conceptual schematic visualization of the currently complex and redundant dependencies between stakeholders during the development of multiple personalized serious games. (B) Schematic visualization of decoupling these dependencies by implementing reusable components that can be used for multiple serious games, thereby removing the tedious and repetitive effort of the stakeholders.

Methods

Use and Design of the Software Engineering Framework

The proposed framework provides a software engineering perspective on how the design process of serious games can be made more efficient by focusing on personalization and reusability. It can be used to transform existing (serious) games into personalized serious games as well as when designing new personalized serious games. As the proposed framework is a software engineering framework, it acts as an intermediate step between conceptualization and implementation, thereby aiming to close the gap between serious game design and implementation.

The Mechanics, Dynamics, and Aesthetics framework is widely accepted in game design as a formal approach for conceptualizing the dynamic behavior of game systems [63]. It approaches a game from the perspective of the player and discerns mechanics (ie, the actions, goals, and rules of the game), dynamics (ie, the behavior followed by the player’s interaction with the mechanics), and the aesthetics (ie, the desired emotional responses of the player when playing the game). The Mechanics, Dynamics, Aesthetics, and Outcomes (MDAO) framework is an extension of the Mechanics,
Dynamics, and Aesthetics framework for the conceptualization of serious games [64], and it introduces the concept of outcomes, that is, the behavioral or intellectual responses of the player after playing the game. The MDAO framework approaches a game from the player’s perspective by first defining the outcomes and aesthetics, followed by dynamics and mechanics. To define the necessary concepts, a domain expert (outcomes and aesthetics) and a game developer (dynamics and mechanics) are needed. The results of the MDAO framework can then be used in the process of answering the 3 questions the framework poses: Why is the game personalized? What parameters are used for personalization? and How is personalization achieved?

Furthermore, the proposed framework is an extended version of the adaptive experience engine proposed by Bellotti et al [65]. They proposed an architecture that decouples the content of educational tasks from the game aspects in sandbox serious games, that is, games that encourage free play, to standardize the development of educational serious games and increase the efficiency by focusing on the reusability of educational tasks. The adaptive experience engine of Bellotti et al [65] uses educational content or tasks, defined by pedagogical experts, and stores these tasks in a common repository for reuse in different games. A game author is then responsible for specifying the requirements of the delivery of such a task at run time, that is, determining the type of task and time it is relevant during gameplay.

We proposed an extension of this framework for the personalization of serious games for health (SGH). We defined the necessary building blocks to increase the reusability of the different components and assign stakeholder responsibilities to these components. Instead of a repository of educational tasks, a knowledge base, which was designed according to the requirements of the domain experts, was included to model the necessary domain knowledge of a serious game. Furthermore, a personalization engine allows for the inclusion and evaluation of multiple personalization algorithms. The proposed framework was evaluated by implementing a proof-of-concept serious game to validate that the framework meets the expectations of the authors.

**Generic Framework**

**Overview**

The framework decoupled the domain knowledge from the personalizable variables in the serious game and personalization algorithms. This allowed the modification or addition of expert knowledge and game concepts independently of each other and the exploration of different personalization strategies without altering the structure of the game itself. The generic architecture discerned 6 modules through which a personalization loop flowed, moving through the different modules to translate user and game data into a specific knowledgeable game task that resulted in the adaptation of the game, as shown in Figure 3.

![Figure 3. The generic framework consists of 3 types of modules. The Knowledge Base module formalizes the knowledge of the domain expert. The game-specific modules are the responsibility of the (game) developer and contain the personalized parameters. The final module, the Independent Personalizer, is the responsibility of the software engineer, who implements the algorithms for personalization.](https://games.jmir.org/2023/1/e40054/fig3.png)

**Domain Knowledge**

The Knowledge Base module is at the heart of the framework and aims to answer the question, Why is there a need for personalization? More specifically, this module is responsible for modeling the knowledge of the domain experts that gives insight into the serious aspect of the game. Serious games are defined as games that have objectives other than pure entertainment. For example, in the case of a rehabilitation scenario, this component will contain certain information regarding rehabilitation exercises and the conditions under which they can be executed, for example, injury type, skill level, and rehabilitation progress. For this example, the expert knowledge would indicate that personalized support during rehabilitation is necessary to accommodate different injury types, different phases of the rehabilitation process, skill levels,
or more, thereby answering why there is a need for personalization. Such expert knowledge can be modeled or computerized using different approaches, for example, simple approaches such as databases or functions or more complex constructs such as ontologies. The goal of the Knowledge Base module is to allow the transferability of knowledge, and the module can thus be reused or replaced based on the domain in which the serious game is situated.

**Game-Specific Modules**

The 3 modules are game specific, namely, the Game and User module, which contains the game itself, and the Feature Abstraction and Action Abstraction modules, which function as a layer of abstraction between the game and the rest of the system. In consultation with the domain expert, the (game) developer can identify the necessary parameters for personalization, answering the question, What needs to be personalized? Two sets of parameters, namely, features and actions, are discerned at design time. First, the information that triggers personalization is defined as a feature. These features can be game information, such as scores, or physiological information collected using wearables (eg, heart rate). These features, which trigger the personalization, introduce a feedback loop in the system to ensure that it is adapted according to the needs of the user. For example, when a score or heart rate is too high, the game might be too exhausting, which should result in the adaptation of the system. In addition, when the user continues to score badly, the game might be too difficult, which in turn will result in the adaptation of the difficulty level. Second, the actions are a set of game parameters that are personalizable based on the context information, such as speed and difficulty level, or more complex constructs, for example, to personalize a storyline. At run time, the features are then periodically sent to the Feature Abstraction module, where they are abstracted to a generic format that can be interpreted by the game-independent modules. Existing player-type frameworks, such as the Hexad Framework, can be used to identify the parameters for personalization [64].

**Independent Personalizer**

Two modules, namely, the Interpreter module and the Personalization Engine, are game and domain independent and can thus be replaced or reused for the personalization of new or existing games. The software engineer aims to answer the question, How is personalization achieved? in these modules without the need to interact with the game itself. The Interpreter module interacts with the Knowledge Base and the game-specific modules to fetch the necessary information to understand the data it has just received. The module then interprets and translates the data into a format that can be understood using the Personalization Engine. This value contains the necessary information for the Personalization Engine to determine whether personalization is necessary and the degree of personalization without the need for context information. For example, if the user has an extremely high heart rate in a relaxation game, the Interpreter will understand, based on the context and domain knowledge that it receives, that the user's heart rate needs to be lowered and personalization has to be applied accordingly. Next, the Personalization Engine, containing one or more models for personalization, applies the action and sends it to the Action Abstraction module, a game-specific module that knows the game task that this action maps (eg, changing the speed of the game). In turn, the Action Abstraction module sends a game-specific task order to the game, which can then adapt accordingly and complete the personalization loop.

Different approaches for personalization exist, for example, intelligent algorithms such as reinforcement learning or recommender systems. By decoupling the personalization task from the game and domain knowledge, opportunities to explore the utility of different personalization models arise. Within one game, different approaches can be used to process or compare triggers for personalization. Furthermore, existing models can be reused for the development of new serious games, which simplifies and reduces the cost of creating a serious game.

**A Serious Game for Shoulder Rehabilitation**

In physical rehabilitation, where patients need to repeat exercises regularly to train their mobility or balance, serious games can provide a welcome distraction from the repetitiveness of the treatment. Patients often lack the incentive to complete the time-consuming exercises at home or start rushing through, resulting in a decline in progress and exercise completion and once again demotivating the patient to correctly adhere to the treatment. Serious games for physical rehabilitation aim to motivate patients to increase their treatment adherence and effectiveness.

As a proof of concept, an existing game was transformed into a serious game and then used to illustrate how the proposed framework can be used to personalize existing serious games as well as when designing new serious games. The game was simple and contained a character that could be controlled by pressing a single button to make the character move upward and avoid upcoming obstacles, similar to the well-known game Flappy Bird (Gears). For this research, the game was transformed into a shoulder rehabilitation exercise that was suitable for physical rehabilitation after injury. As mentioned previously, in physical rehabilitation, it is important for patients to perform their exercises regularly and correctly.

The mechanics of the chosen game set the users up for failure, as they must continue controlling the character until it hit an obstacle and the game was over. This seems contradictory, as rehabilitation patients should not be rushed when performing their exercises, and this thus indicates the need for personalization by adapting the difficulty of the game to the capabilities of the user. Instead of controlling the character with a button, the user could control it by lifting their outstretched arm to the shoulder level and moving the character upward. The Intel RealSense Camera (Intel) [66] and the Cubemos skeleton (Intel) tracking SDK [67] were used to track the movement of the arm of the user and control the upward movement of the bird, as shown in Figure 4. The user earned a score based on how long they manage to keep the bird in the air without hitting any obstacles; otherwise, the game ended.
Proof of Concept: an Adaptive and Personalized Serious Game

The following paragraphs discuss the different components of the generic framework in detail, using the implementation of the serious game for shoulder rehabilitation as an illustration of how the framework works. Figure 5 provides an overview of the implemented proof of concept using the proposed framework.

Domain Knowledge

As mentioned in the section Generic Framework, the Knowledge Base models the knowledge that is relevant for the specific serious game. This can be achieved using a simple database, storing values, or more complex constructs such as ontologies that are capable of modeling complex relationships between concepts in a computer-readable format. The responsibilities of the domain expert consist of (1) defining or reusing the necessary domain knowledge and, together with the (game) developer, (2) identifying the personalizable parameters of the serious game, that is, features and actions.

For this proof of concept, the Knowledge Base was kept simple. The Knowledge Base stored 2 functions, one for each identified feature, namely, $f_{\text{heart\_rate}}$ and $f_{\text{score}}$. The function $f_{\text{heart\_rate}}$ calculated the highest accepted heart rate for the given context information, whereas the function $f_{\text{score}}$ looked at previous $N$ scores, given the necessary context to evaluate the performance of the user. This means that, based on the game and its objective, 2 features were identified, namely, heart rate and game score.

Game-Specific Modules

As previously mentioned, for the game, 2 features were identified, namely, heart rate and game score. These data were collected by the Game and User component and were used to trigger the start of the personalization loop as they were periodically sent to the Feature Abstraction module, which sent them to the independent personalization modules for evaluation. As patients cannot rush their rehabilitation exercises but should perform them correctly to increase their mobility, the speed of the game should be adjusted accordingly. This is achieved by the Personalization Engine, which adapts the game speed based on the user’s performance.
the game was identified as a personalizable action. If the heart rate of the user increased or their game score decreased, the speed of the game should be lowered to continue to ensure good shoulder exercise performance.

Personalization of the game can occur both online (ie, in real time) and offline. For offline personalization, the feature data were only sent after a gaming session, adapting the game for the next session, whereas for real-time personalization, the game was personalized during the gaming session, based on the data received up to that point. One game session was considered to be a level of the game. Both types of personalization have their benefits: real-time personalization allows the game to quickly respond to the user’s currently changed context, whereas offline personalization facilitates a more complex analysis that considers the overall performance and progress of the user, instead of just a moment in time.

For each feature, the Feature Abstraction module contains its respective Transformer module, responsible for translating the data to a generic format, which can be interpreted by the independent personalization modules. Using this generic format, the features can be interpreted by the rest of the system without the need for game-specific information. Nonetheless, these features had very little meaning without the necessary context information. For example, it was difficult to interpret a heart rate of 100 without any additional information. However, given the knowledge that this was the resting heart rate of a 25-year-old male, it was possible to interpret the importance if this heart rate. Therefore, the Context Locator module was responsible for providing the relevant context information to the personalization modules when needed. The Context Locator knew which context mapped to what features and fetched this information from the Game and User module.

This decoupling reduced the integration of a new serious game for personalization to the following steps: (1) identifying the features, (2) identifying the actions, (3) implementing or reusing the respective Transformer modules, and (4) creating a Context Locator that fetched the relevant context information for the correct feature.

**Independent Personalizer**

After the Interpreter receives the feature data in a generic format, it contacts the Knowledge Base to fetch the information to interpret the received value. This information includes the expert knowledge and context parameters that are linked to a specific feature. For the game, the knowledge linked to the heart rate feature was \( f_{\text{heart_rate}} \), and the relevant context parameters were age, sex, and game intensity. For the score feature, the linked knowledge was \( f_{\text{score}} \) and the relevant context parameters were the previous N scores.

Next, the Interpreter module contacts the Context Locator to fetch the values of these context parameters for each feature. After receiving all the required information, the Interpreter can give meaning to the feature data and transform it into a processed value that contains the necessary information for the Personalization Engine, that is, the degree of the needed personalization, but is devoid of any game- or domain-specific knowledge. Continuing the example, the feature data, a heart rate of 100, and the associated context, the resting heart rate of a 25-year-old male, can be interpreted as an unusually high heart rate that should be reduced.

The Personalization Engine receives this processed value, and its Model Locator sends it to the correct model. As mentioned in the previous section, multiple types of models exist; however, for the proof of concept, only simple rule-based models were implemented to illustrate the framework. The first model receives the processed score as input and is used as an offline adaptation model, that is, the score after each gaming session is processed and used for the adaptation of the speed for the next gaming session. The second model takes the processed score and heart rate as input and is used for offline adaptation, whereas the third model receives the same input but is used for real-time adaptation.

The output of these models is a personalized value that is sent to the Action Abstraction module. This module, more specifically, the Speed Transformer, performs the reversed action of the Feature Abstraction module, as it transforms this generic format into a game-specific format, namely, to an action of the game, that is, speed.

As the Interpreter processes these input values, and the Action Abstraction module processes the output values, the software engineer does not have to have any domain- or game-specific knowledge to implement personalization algorithms. This reduces the responsibilities of the software engineer in (1) implementing the Interpreter module and (2) developing or reusing personalization models. Figure 6 presents an overview of the responsibilities of each of the involved stakeholder.
Results

Simulation of Game Scores

To evaluate the proposed framework, the implementation of a personalized game is assessed using several simulations. First, the response of the system was evaluated when only game scores were used as features. Each score represents the score obtained after a single game. The difficulty level at the start of a gaming session was always 0.8. Adaptation of the speed would only occur after a completed game, that is, offline personalization.

The first series of scores was simulated to show a near-horizontal trend; that is, the user achieved, on average, a constant score for each game. The system used the previous scores of the user to decide whether the difficulty, that is, the speed of the bird, must be personalized. Thus, the difficulty was expected to be increased if this trend continued. Figure 7 shows the simulated scores that were fed to the system (top panel) and the response of the system regarding the adaptation of the difficulty (bottom panel). The gaming scores from the first 10 games showed a more diverging scoring pattern, indicating some drops in the score, and the system responds accordingly by lowering the difficulty. Game 14 showed a substantial increase in the gaming score, followed by a nearly constant score. As shown in the figure, the system reacted to this trend as expected by continuously increasing the difficulty of the game.

A second simulation showed continuously improving game scores, that is, on average, upward trend, as illustrated in Figure 8. As the user continued to improve their previous score, the game was most likely to become too easy for the user. The expected response of the system was to increase the difficulty of the game more visibly during this upward trend than during a nearly constant trend. As of game 16, after a brief drop, the user significantly improved their game scores, upholding an upward trend. As expected, the system countered this by increasing the difficulty significantly faster than during the first 13 sessions, when the user achieved a constant score.

The final simulated game series illustrated the situation in which the performance of the user continued to drop, that is, a downward trend in the game scores, as illustrated in Figure 9. Such behavior might indicate that the game is too difficult for the user and the game should respond by significantly decreasing the difficulty. As of the 16th game, the score of the user kept dropping, to which the system responded, as expected, by decreasing the speed of the game. By game 27, the user reached a new local maximum, which was answered by the system by again a slow increase in difficulty.
Figure 8. The score of the user is, on average, constant during the first 13 games (above), to which the system responds with a slight increase in difficulty (below). After a brief drop in scores, the user’s score indicates an upward trend as of game 16. The system responds by significantly increasing the difficulty of the game as long as this upward scoring trend continues.

Figure 9. In the first 15 games, the user reaches, on average, a constant score (above), to which the system again responds by slowly increasing the difficulty (below). As of game 16, the score keeps dropping, showing a downward trend, to which the system responds by significantly lowering the difficulty.

Simulation of Game Scores and Heart Rate

For a second evaluation of the system, game scores and a rising heart rate, which at some point would exceed the set maximum heart rate of 180, were simulated. For this particular use case, where the user should perform a simple arm exercise, an elevated heart rate could indicate that the patient was distressed or experiencing pain, which would negatively affect the obtainable goal, that is, increasing the mobility of the arm. Therefore, the system should respond to this elevated heart rate by adapting to the game. The response of the system was compared for both offline, that is, after a gaming session, and real-time personalization, that is, during a gaming session.

For this simulation, we expected that a high heart rate would have a greater impact on the personalization of the game than the gaming score. If a rising heart rate was detected and the heart rate approached the maximum threshold heart rate of 180, the system should respond by immediately reducing the speed of the game. However, for offline personalization, the system was only able to respond if the average heart rate of the last completed game session exceeded 180. Therefore, the system could only reduce the speed of the game after a completed game session. For real-time personalization, the system was expected to respond much sooner and decrease the speed of the game from the moment the real-time heart rate exceeds the threshold value during the gaming session.

For this evaluation, the constant gaming score from the first simulation (Figure 7) was reused. Figure 10 illustrates the simulated gaming score (above), simulated average heart rate per gaming session (middle), and response of the system for both offline adaptation and real-time adaptation (below). For offline adaptation (green), the difficulty during a gaming session remained constant and was adapted only after the game has been completed. For real-time adaptation (purple), the difficulty could vary during a gaming session as real-time adaptations occurred. The difficulty values illustrated in Figure 10 correspond to the difficulty level at the end of each gaming session.

In the first simulation of a constant game score, the expected behavior was an increase in the difficulty. However, the system then also had to consider a rising heart rate, which resulted in more or less constant difficulty as long as the heart rate continued to rise. For offline adaptation, as soon as the average heart rate of a game session crossed the threshold value, which was the case for game 27, the system responded by decreasing the difficulty. Because offline adaptation introduced a certain delay, the game could only respond by game 28. The figure shows that this delay was not visible in the case of real-time adaptation, as the difficulty had already decreased for game 26. Figure 11 provides a detailed overview of how the system reacted (below) to the real-time heart rate (above) during games 26 to 30. Here, it is clear that the user’s heart rate already
exceeded the threshold value during game 26, and thus the system could immediately reduce the difficulty of the game in the case of real-time adaptation. Because of the rigidity of offline adaptation, that is, one set speed for an entire game, the difficulty also decreased more slowly compared with real-time adaptation. The minimum difficulty for offline personalization of the entire exercise session was reached in game 30, whereas this minimum was already reached in game 27 using real-time personalization, reaching an even lower speed by the end of game 30.

**Figure 10.** For an, on average, constant score (above) and a constantly increasing heart rate (middle), the system responds differently for real-time personalization compared with offline personalization (below). If the system is updated after the game (offline personalization), the maximum threshold heart rate of 180 bpm is exceeded in game 27, of which the average heart rate is 182 bpm. The system thus starts decreasing the difficulty as of game 28. When the system is updated during the game (real-time personalization), the system already decreases the difficulty as of game 26, therefore achieving a much lower speed much faster. bpm: beats per minute.

**Figure 11.** This detailed overview of the heart rate of the user starting from game 26 (top) indicates that the system can reduce the difficulty of the game (bottom) much faster in the case of real-time adaptation as the maximum threshold of 180 is already exceeded in game 26. Because the average heart rate of a gaming session only exceeds 180 in game 27, a delay is introduced in the version using offline adaptation. Real-time adaptation, therefore, allows the system to respond much faster to critical values than when offline personalization is used.

**Discussion**

**Principal Findings**

This study proposes a framework to standardize and simplify the design of personalized SGH. The process of designing a serious game is challenging and involves multiple stakeholders. The proposed framework identifies the responsibilities of the involved stakeholders using 3 key questions. The framework focuses on the reusability, formalization, and transferability of the expertise of these stakeholders to simplify this process. As a proof of concept, a simple game was transformed into a serious game for shoulder rehabilitation. The game was implemented according to the proposed framework to illustrate how it can streamline the design and implementation of personalized serious games. Several simulations were conducted to evaluate the personalization mechanics of the resulting personalized serious game.

The integration of the framework introduces several advantages for the design of personalized serious games. First, by assigning responsibilities to the stakeholders involved, the cocreation of the serious game is still respected, but the complicated process is simplified. Each stakeholder knows what is expected of them and with whom they need to communicate to receive specific information. By introducing 3 types of modules, namely, domain
knowledge, game-specific modules, and an Independent Personalizer, and defining the communication between these modules, the reusability of these modules becomes possible, introducing a sort of “plug-and-play” structure of the components. Domain knowledge can be formalized using complex structures, such as ontologies or a simple information database, and can be reused or extended to multiple serious games. Over time, researchers can gather an extensive knowledge base covering one or more domains, each time reducing the effort needed to integrate their knowledge into a new serious game. A similar approach can be adopted for the other modules.

Furthermore, this framework allows for the integration of multiple personalization algorithms or approaches. As discussed in the Introduction section, different methods exist for the personalization of serious games, but very little is known about their effectiveness. Using this framework, software engineers can focus on further developing and evaluating these personalization approaches without the need to repeat the entire process from scratch. Thus, more effort can be directed toward developing more complex and intelligent algorithms. However, the framework allows not only the reusability of algorithms but also multiple algorithms to be used simultaneously for one serious game, as the Personalization Engine can contain multiple models for the identified features. This can be interesting when both real-time and offline personalization needs to be considered. Serious games in the health care domain often deal with gathering time-sensitive information, collecting health sensor data, or monitoring the performance of the user, and wrong movements can have detrimental effects. The framework allows the integration of specialized personalization models that monitor these data and intervene instantaneously when needed without the need to interrupt the rest of the personalized game mechanics, as these models can exist independently of one another.

Finally, the literature has shown that the evaluation of the effectiveness of serious games is often inadequate because of sparse evaluation results, lack of standardized evaluation approaches, and high cost of designing serious games. This framework takes the first steps toward a more accessible and standardized evaluation procedure for serious games. The framework easily allows the control of different parameters by interchanging the components under evaluation, such as personalization algorithms, and reusing other components. Moreover, because of the reusability of components, different personalization strategies for serious games can be tested and evaluated at a much faster pace.

Future Work

Although this study offers many contributions, it has some limitations that will be addressed in future work. First, the complexity of the proof of concept is limited to illustrating the functioning of the framework. Second, the resulting personalized serious game was evaluated using simulations. Therefore, future work will focus on implementing a more complex knowledge base and intelligent algorithms. Using this framework will facilitate the evaluation and comparison of different artificial intelligence algorithms, such as recommender systems or machine learning algorithms. The resulting serious game will be assessed using large-scale evaluations with end users to ensure that the personalization strategies used fit the needs of the target audience. To this end, questionnaires such as the System Usability Scale [68] and the Gameplay Scale [69] will be essential for thorough evaluation.

Conclusions

We proposed a framework for the design of personalized SGH. The aim of this framework is to offer guidelines that streamline the complex co-creation process of serious games. Moreover, the implementation of this framework facilitates the transferability of domain knowledge and reusability of personalization algorithms, thereby taking the first steps to the state of the art concerning personalization in SGH. Complex models of specific domain knowledge can be reused and extended, while different personalization strategies can be easily compared and evaluated.

Authors’ Contributions

SC, VN, and FDB conceptualized the study. SC and VN designed the methodology; were responsible for validation, formal analysis, investigation, and curation; and wrote the original draft. VN procured the software. SC visualized the data. SC, VN, FDB, and FDT reviewed and edited the manuscript. FDT supervised the study. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References


Vertical Integration of Electronic Health Records in Medical Consortiums: Dynamic Modeling Approach Based on the Evolutionary Game Theory

Shenghu Tian¹,², PhD; Yu Chen², PhD

¹Business School, Yunnan University of Finance and Economics, Kunming, China
²School of Management and Economics, Kunming University of Science and Technology, Kunming, China

Corresponding Author:
Yu Chen, PhD
School of Management and Economics
Kunming University of Science and Technology
Number 727, Jingming South Road, Chenggong District
Kunming, 650504
China
Phone: 86 0871 6517 ext 2836
Email: Ychen@kust.edu.cn

Abstract

Background: China has continuously issued policies to speed up the interconnection, mutual recognition, sharing of medical information systems, and data integration management across regions and institutions. However, the vertical integration of electronic health records (EHRs) within the medical consortium is hampered by “poor mechanism and insufficient motivation” and the phenomenon of “free riding” among participating medical institutions, which makes the integration less effective.

Objective: We hope to clarify the game mechanism of stakeholders in the vertical integration of EHRs, and put forward targeted policy suggestions for improvement.

Methods: We constructed the “government-hospital-patient” tripartite evolutionary game model based on the detailed analysis of the research problems and their assumptions. We then simulated the game strategies and outcomes of each participant using the system dynamics approach to reveal the long-term strategy evolution mechanism of the core participants in the vertical integration of EHRs in the medical consortium, as well as the influencing factors and action mechanisms of each party’s strategy evolution to provide references for improving relevant policies.

Results: The evolutionary game system could eventually reach an optimal equilibrium, but in areas where the government was required to be in a dominant position, patient supervision was necessary to have a positive role, while a reasonable reward and punishment mechanism can promote active participation of hospitals.

Conclusions: The effective way to achieve the goal of vertical integration of EHRs in the medical consortium is to build a multiagent coordination mechanism under the guidance of the government. Meanwhile, it is necessary to establish a scientific integration performance evaluation mechanism, a reward and punishment mechanism, and a benefit distribution mechanism to promote the healthy development of vertical integration of EHRs in medical consortiums.

(JMIR Serious Games 2023;11:e41528) doi:10.2196/41528

KEYWORDS
medical consortium; EHRs; evolutionary game; system dynamics; medical information resources

Introduction

The construction of a medical consortium is an important institutional innovation to deepen medical reform and a specific practical path to achieve graded diagnosis and treatment, which plays a vital role in the reform and development of medical and health undertakings in China. It is generally believed that a medical consortium is a medical service community with high-level medical institutions (tertiary hospitals) as the core institution (also named the leading unit) and other medical and health institutions at various levels in the region united as participating units (also known as member units) to vertically integrate medical and health resources, thereby eliminating
fragmentation of medical services and achieving integration of services, responsibilities, benefits, and management [1]. The accessibility, continuity, and radiation of medical services are enhanced, and the quality and efficiency are improved, via the integration of cross-regional and cross-level medical services, resources, technologies, and information [2]. Since the promulgation of the “Pilot Work Plan for the Construction of Urban Medical Consortium” [3], relevant policies have repeatedly emphasized strengthening the integrated management of medical consortiums, realizing the continuous recording of electronic health records (EHRs) and electronic medical records, and gradually realizing the effective sharing of medical and health information within the medical consortium, to improve the efficiency of treatment and reduce medical costs. The “14th Five-Year Plan for the Implementation of Quality and Efficient Medical and Health Services,” released in July 2021 [4], also includes accelerating the interconnection, mutual recognition, and sharing of cross-regional and cross-institutional information systems, as well as data integration and management as key projects. The construction of a medical consortium has thus become an important strategic measure to build an integrated medical and health service system in China.

EHRs refer to the systematic digital record of health information, such as disease treatment, immunization, chronic disease management, genetic history, throughout the life cycle of individual citizens through a variety of collection channels. The existing EHRs are scattered in different medical and health institutions at all levels, which seriously restricts the healthy and orderly development of a medical consortium. The vertical integration of EHRs refers to the collection, sorting, and ordering of EHRs among medical institutions at different levels within a medical consortium [5]. The purpose of EHRs is to break the decentralized pattern, promote the collaboration and hierarchical diagnosis and treatment within the medical consortium, promote the in-depth cooperation and organizational collaboration within the medical consortium, and help the development of the medical consortium from loose to compact. The integration of medical information resources has always been the core of medical and health information research. Although existing studies focused on the integration of diagnosis and treatment information resources within medical institutions, attention given to the integration and sharing of EHRs across institutions is far from adequate. The reasons for this are as follows: (1) the policies related to the integration of medical consortium and EHRs have not been proposed for a long time, and the policy effect has not been fully manifested; and (2) the shackles of “poor mechanism and insufficient motivation” are deeply rooted in the integration of medical resources across institutions, and the phenomenon of “free riding” by participating medical institutions is prominent, which makes the integration ineffective.

The game theory is suitable for the study of multiagent behavior. Evolutionary game is favored because its bounded rationality assumption of game players is more realistic, which is widely used in various fields, such as multiple governance, organizational coordination, and information behavior. Cross-institutional information resource sharing often involves many stakeholders, and it could gradually approach the evolutionary stable state after repeated games. Some studies have tried to use the evolutionary game method to perform relevant research in the medical and health field, such as studying doctor-patient disputes under government regulation [6], evolutionary game of a 2-way referral mechanism [7], emergency management of public emergencies [8], knowledge sharing in online health communities [9], drug safety and public health quality supervision [10,11], and integrated health care systems [12]. Some studies also use evolutionary game and simulation methods to explore the willingness of medical data sharing, but these only consider the 2 parties of patients and medical service institutions [13], and fail to consider the willingness and influence mechanism of cross-agency data sharing of the government and other multiparties under the organization mode of the medical consortium. The government, hospitals, and patients are the 3 direct stakeholders involved in the vertical integration of EHRs. To make all parties act in a unified manner in order to achieve the collective goal of resource integration, it is necessary to understand the long-term game strategy and its evolutionary stability of all parties. In addition, most previous studies used qualitative analysis methods such as speculative reasoning, expert interviews, or text analysis, and few studies used mathematical methods to discuss the issues shared by EHRs across institutions.

The essence of cross-organizational EHRs integration within a medical consortium is not a technical implementation but an organizational synergy problem, the core of which is the progressive evolutionary game problem of multiple stakeholders. In particular, participating medical institutions and patients have more evolutionary characteristics of progressive learning, and government behavior affects the development direction of EHRs integration. In this paper, we used evolutionary game analysis to incorporate the government, hospitals, and patients into a system model to explore the long-term strategy evolution of the 3 parties in the vertical integration of cross-organizational EHRs within a medical consortium, as well as the influencing factors and mechanisms of each party’s strategy evolution. Finally, we proposed corresponding countermeasures based on data simulation using system dynamics. The findings of the study can provide a reference basis for the smooth implementation of vertical integration of EHRs within a medical consortium.

Methods

Construction of an Evolutionary Game Model

Problem Statements and Assumptions

From the perspective of stakeholders, the core participants involved in the vertical integration of EHRs are mainly the government, hospitals, and patients. As the main driving force of the hierarchical diagnosis and treatment system, the government has made every effort to promote the vertical integration of EHRs in the medical consortium through top-level design and the introduction of relevant policies to optimize the allocation of medical resources. However, local governments may make every effort to promote it or treat it negatively, due to the impact of protectionism, high costs, human resources, and complexity of implementation. Therefore, the behavior
strategies of the local government are either “trying to promote it” or “treating it negatively.”

**Hypothesis 1**

For local governments, if they try their best to promote the integration of EHRs, the income, such as social value, public praise, superior rewards, will be $R_g$ and the cost will be $C_g$; when the local government takes a negative attitude, it will neither give subsidies or rewards to the hospitals and patients nor impose penalties. The cost is $K_g(C > K_g)$ in such cases, and the local government will be punished as $F_g$ (including the accountability of the superior government, government credibility, and reputation loss). When the hospital actively participates in the information integration under the policy effect and achieves good results in practice, the boundary between the local government’s “efforts to promote” and “negative treatment” will become very vague. Therefore, the local government will not be held accountable by the superior government, nor will there be potential punishment such as loss of credibility and reputation.

As the owner of EHRs, hospitals’ participation attitude directly affects the success or failure of the vertical integration of EHRs. Existing research shows that the leading large hospitals worry about weakening the absolute dominant position of medical resources, while the small hospitals participating in the medical consortium may negatively treat the integration of information resources to maintain their own economic interests and social reputation [14]. However, both the leading hospitals and member units may actively respond to national policies, abandon “free riding” opportunism, and actively participate in the vertical integration project of EHRs. Therefore, the behavior strategies of the hospital are active participation or passive participation.

**Hypothesis 2**

For the hospital, the revenue under normal operation is $R_h$. If the hospital actively participates in the sharing and utilization of cross-institutional EHRs, the reward subsidy from the superior government is $F_h$, and the cost is $C_h$. If the hospital chooses a negative strategy, its cost is $K_h(C > K_h)$, and the resulting damage to the overall social interests is $R_p$, which will be borne by the public (especially patients). When patients find that their EHRs cannot be shared and utilized across institutions, if they complain to the superior supervision organization, the hospital will be punished as $M_p$. If the patients do not choose to complain, no supervision cost will be incurred.

As the biggest beneficiary of the hierarchical diagnosis and treatment system, the patients have the right and obligation to supervise and complain about the interconnection and availability of EHRs in the medical consortium. The patients may supervise hospital behaviors to protect their own rights and interests and complain to the medical and health supervision department, but they may also give up their right to supervision because of negative ideas such as “It’s better to save trouble” and “supervision is the same as non-supervision” or because they think that the supervision cost is too high. Therefore, the behavior strategy space of patients is participating in supervision or giving up supervision.

**Hypothesis 3**

For patients, the cost of active participation in supervision is $C_p$, and the reward from the government is $F_p$; if patients give up supervision, they will neither incur costs nor receive government incentives.

The behaviors of local governments, hospitals, and patients all conform to the bounded rationality hypothesis of the evolutionary game theory. In the game process, each party continuously adjusts its own strategies through the behaviors of the other 2 parties, and finally reaches a stable state after repeated games.

**Construction of a Tripartite Profit Matrix**

Assuming that the ratio of local government choosing to promote is $x$, the ratio of its negative treatment is $1 - x$. If the rate of active participation is $y$, the rate of passive participation is $1 - y$. If the rate of patients choosing supervision complaints is $z$, the rate of giving up supervision is $1 - z$, and $x, y, z \in [0, 1]$. The relevant parameter settings and their meanings are shown as Table 1.

According to the aforementioned analysis and assumptions, the game payoff matrix of the 3 core players in the vertical integration of EHRs in the medical consortium under different strategic combinations is shown in Table 2.
Table 1. Relevant parameters setting and their meanings.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_g$</td>
<td>Benefits when the local government tries to promote EHR\textsuperscript{a} integration</td>
</tr>
<tr>
<td>$C_g$</td>
<td>Costs when the local government pushes hard</td>
</tr>
<tr>
<td>$K_g$</td>
<td>Costs when the local government treats EHR integration negatively</td>
</tr>
<tr>
<td>$F_g$</td>
<td>Punishment received when the local government treats EHR integration negatively</td>
</tr>
<tr>
<td>$R_h$</td>
<td>Revenue when the hospital is in normal operation</td>
</tr>
<tr>
<td>$C_h$</td>
<td>Costs when hospitals are actively involved in EHR integration</td>
</tr>
<tr>
<td>$K_h$</td>
<td>Costs when hospitals are passively involved in EHR integration</td>
</tr>
<tr>
<td>$F_h$</td>
<td>Government subsidies when hospitals actively participate in EHR integration</td>
</tr>
<tr>
<td>$M_h$</td>
<td>Punishment when the hospital participates passively in EHR integration</td>
</tr>
<tr>
<td>$F_p$</td>
<td>Rewards received when patients actively participate in supervision</td>
</tr>
<tr>
<td>$C_p$</td>
<td>Costs when patients actively participate in supervision</td>
</tr>
<tr>
<td>$R_p$</td>
<td>Social damage caused by negative hospital participation</td>
</tr>
</tbody>
</table>

\textsuperscript{a}EHR: electronic health record.

Table 2. Evolutionary game payoff matrix of the local government, hospital, and patient.

<table>
<thead>
<tr>
<th>Game players</th>
<th>Local governments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Push hard ($x$)</td>
</tr>
<tr>
<td>Hospital</td>
<td></td>
</tr>
</tbody>
</table>

**Actively participation ($y$)**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation in supervision ($z$)</td>
<td>$R_g - C_g - F_h - F_p, R_h - C_h + F_h, F_p - C_p$</td>
<td>$-K_g, R_h - C_h, -C_p$</td>
</tr>
<tr>
<td>Waiver of supervision ($1-z$)</td>
<td>$R_g - C_g - F_h, R_h - C_h + F_h, 0$</td>
<td>$-K_g, R_h - C_h, 0$</td>
</tr>
</tbody>
</table>

**Passive participation ($1-y$)**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation in supervision ($z$)</td>
<td>$R_g - C_g + M_h, -K_h - M_h, F_p - C_p - R_p$</td>
<td>$-K_g - F_g, -K_h - C_p$</td>
</tr>
<tr>
<td>Waiver of supervision ($1-z$)</td>
<td>$R_g - C_g + M_h, -K_h - M_h, -R_p$</td>
<td>$-K_g - F_g, -K_h - R_p$</td>
</tr>
</tbody>
</table>

**Behavior Equilibrium Analysis of Evolutionary Game**

**Expected Profit of Game Players**

Let $E_{h1}$ be the expected return when the hospital actively participates, $E_{h2}$ be the expected benefit when the hospital passively participates, and $E_r$ be the average benefit in both cases of active and passive hospital participation, then we have:

$$E_{h1} = yE_{h1} + (1-y)E_{h2}$$  \hspace{1cm} (4)

$$E_{h1} = xz(R_g - C_h + F_g) + x(1-z)(R_g - C_g - F_h) + z(1-x)(R_g - C_g - F_h) + z(1-x)(1-z)(R_h - C_h)$$  \hspace{1cm} (5)

$$E_{h2} = xz(-K_h - M_h) + x(1-z)(-K_g - M_g) + z(1-x)(-K_h - M_h) + z(1-x)(1-z)(-K_h - M_h)$$  \hspace{1cm} (6)

Let $E_{p1}$ be the expected benefit when the hospital actively participates, $E_{p2}$ be the expected benefit when the hospital passively participates, and $E_p$ be the average benefit in both cases of active and passive hospital participation, then we have:

$$E_{p1} = zE_{p1} + (1-z)E_{p2}$$  \hspace{1cm} (7)
Combining equations 1-3, we could construct the replication dynamic equation when the government tries to promote EHR integration as follows:

\[
F(x) = dx/dt = x(E_{g1} - E_{g2}) = x(1-x)(E_{g2} - E_{g1})
\]

\[
= x(1-x)\{[yz(R_g - C_g - F_h - F_p) + y(1-z)(R_g - C_g - F_h) + z(1-y)(R_h - C_h + M_h) + (1-y)(1-z)(R_g - C_g - F_h) + M_h] - [yz(-K_g) + y(1-z)(-K_g) + z(1-y)(-K_h) - F_h] + (1-y)(1-z)(-K_g - F_h)]
\]

\[
= x(1-x)[(R_g + F_h + M_h + K_h - C_g - y(z(F_p + F_h + F_g + M_h)] (10)
\]

Combining equations 4-6, the replication dynamic equation for active hospital participation could be constructed as follows:

\[
F(y) = dy/dt = y(E_{h1} - E_{h2}) = y(1-y)(E_{h2} - E_{h1})
\]

\[
= y(1-y)\{[xz(R_h - C_h + F_h) + x(1-z)(R_h - C_h + F_h) + z(1-x)(R_h - C_h) + (1-x)(1-z)(R_h - C_h)] - [xz(-K_h - M_h) + x(1-z)(-K_h - M_h) + z(1-x)(-K_h)]
\]

\[
= y(1-y)[(R_h + K_h - C_h - x(F_h + M_h)] (11)
\]

Similarly, combining equations 7-9, the replication dynamic equation could be constructed when the patient is actively supervised as follows:

\[
F(z) = dz/dt = z(E_{p1} - E_{p2}) = z(1-z)(E_{p2} - E_{p1})
\]

\[
= z(1-z)\{[xy(F_p - C_p) + y(1-x)(-C_p) + x(1-y)(F_p - C_p) + (1-x)(1-y)(-C_p)] - [xy(1-y)(-C_p) + (1-x)(1-y)(-C_p)]
\]

\[
= z(1-z)[x(F_p - R_p - yR_p) + (1-y)R_p - C_p] (12)
\]

**Equilibrium Analysis of the Evolutionary Game System**

Combining equations 10-12, we can get the 3D dynamic system of the game participants (i.e., the government, hospitals, and patients), which is given as follows:

\[
F(x) = dx/dt = x(1-x)[(R_g + F_h + M_h + K_h + C_g - y(z(F_p + F_h + F_g + M_h)]
\]

\[
F(y) = dy/dt = y(1-y)[(R_h + K_h - C_h + x(F_h + M_h)]
\]

\[
F(z) = dz/dt = z(1-z)[x(F_p - R_p - yR_p) + (1-y)R_p - C_p] (13)
\]

By making the 3 replicated dynamic equations in equation (13) equal to 0, we can obtain 8 pure strategies Nash equilibrium solutions of the evolutionary game system, namely, (0, 0, 0), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), and (1, 1, 1). Existing studies in the field have proved that only the stability of the aforementioned 8 local equilibrium points can be considered in the analysis of a 3-party evolutionary game [16], and that it is not necessary to consider the Nash equilibrium solution E (x*, y*, z*) of the mixed strategy. The sign of Jacobian matrix can be used to judge whether the aforesaid 8 local equilibrium points are the evolutionary stability strategy of the system, and the condition is det(J)>0 and tr(J)>0. The Jacobian matrix can be obtained by solving the partial derivatives of the 3 equations in equation (13) with respect to x, y, and z, respectively:

\[
J_{11} = (1 - 2x)(R_g + F_h + M_h + K_h - C_g - y(z(F_p + F_h + F_g + M_h)];
\]

\[
J_{12} = (1 - x)(z(F_p + F_h + M_h)];
\]

\[
J_{13} = (1 - 2y)(R_h + K_h - C_h + x(F_h + M_h));
\]

\[
J_{22} = (1 - 2y)(R_h + K_h - C_h + x(F_h + M_h));
\]

\[
J_{33} = (1 - 2z)(x(F_p + F_h + M_h]);
\]

\[
det(J) = J_{11}J_{22}J_{33} + J_{12}J_{33}J_{13} + J_{13}J_{12}J_{23} - J_{12}J_{13}J_{23} - J_{11}J_{23}J_{32} - J_{12}J_{13}J_{23} - J_{11}J_{22}J_{33};
\]

\[
tr(J) = J_{11} + J_{22} + J_{33} = 1 - 2x + 1 - 2y + 1 - 2z - M_h - C_h + y(F_p + F_h + M_h) + (1 - 2y)(1 - 2z)[C_r + x(F_h + M_h)];
\]

\[
\text{The det}(J) and tr(J) values of 8 local equilibrium points can be calculated from equations 14 and 15 and judgment conditions, as shown in Table 3.}

So far, the evolutionary game system has achieved the stable state under certain conditions. However, it can be seen from Table 3 that the det(J) and tr(J) values are determined by the values of many parameters. The existing conditions and calculation methods make it difficult to determine whether the aforesaid 8 equilibrium points are the evolutionary stability strategy, which indicates that it is not clear whether there are equilibrium points in the evolutionary game system. System dynamics can analyze the complex dynamic evolution process of evolutionary game models under the conditions of bounded rationality and asymmetric information [17], and can explore and reveal the potential mechanism of problems and seek key strategies to solve problems [18]. Therefore, we further use the
system dynamics tool to simulate the impact of uncertainty factors on the stability of the evolutionary game system, so as to provide decision-making reference for formulating relevant policies and optimizing the dynamic mechanism of EHRs integration.

Table 3. The det(\(J\)) and tr(\(J\)) of the Jacobian matrix.

<table>
<thead>
<tr>
<th>Equilibrium point</th>
<th>Det((J))</th>
<th>Tr((J))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 0)</td>
<td>((R_p - C_p)(K_g + R_h - C_l)(F_g + K_g + M_h + R_g - C_g))</td>
<td>(F_g + K_g + M_h + R_h + R_g + R_p - C_g - C_h - C_p)</td>
</tr>
<tr>
<td>(0, 0, 1)</td>
<td>((C_p - R_p)(K_h + R_h - C_l)(F_g + K_g + M_h + R_g - C_g))</td>
<td>(C_p + F_g + K_g + M_h + R_g + R_h - R_p - C_g - C_h)</td>
</tr>
<tr>
<td>(0, 1, 0)</td>
<td>(-C_p(Ch - K_h - R_h)(K_g + R_g + C_g - F_h))</td>
<td>(C_h - C_p - F_h + K_g + M_h + R_g - R_h - C_g)</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>(C_p(C_h - K_h - R_h)(K_g + R_g - C_g - F_h - F_p))</td>
<td>(C_h + C_p - F_h - F_p + K_g + R_h - R_p - C_g)</td>
</tr>
<tr>
<td>(1, 0, 0)</td>
<td>((F_p - C_p)(C_g - F_g - K_g - M_h - R_g)(F_h + K_h + M_h + R_h - C_h))</td>
<td>(C_g - C_h - C_p - F_h + F_p - F_g - K_g + K_h - R_g + R_h)</td>
</tr>
<tr>
<td>(1, 0, 1)</td>
<td>((C_p - F_p)(C_g - F_g - K_g - M_h - R_g)(F_h + K_h + M_h + R_h - C_h))</td>
<td>(C_g - C_h + C_p - F_h + F_p - F_g + K_g + K_h - R_g + R_h)</td>
</tr>
<tr>
<td>(1, 1, 0)</td>
<td>((F_p - C_p)(C_g + F_h - K_g - R_p)(C_h - F_h - K_h - M_h - R_h))</td>
<td>(C_g + C_h - C_p - F_g - F_h - F_p - K_g - K_h - R_g + R_h)</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>((F_p - F_p)(C_g - F_g - F_h + F_p - K_g - R_g)(C_h - F_h - K_h - M_h - R_h))</td>
<td>(C_g + C_h + C_p - K_g + K_h - M_h - R_g + R_h)</td>
</tr>
</tbody>
</table>

Construction of the System Dynamics Simulation Model

According to the aforementioned analysis, we used Vensim PLE (Ventana Systems, Inc) to build a system dynamics model of the evolutionary game among the government, hospitals, and patients in the vertical integration of EHRs in the medical consortium, as shown in Figure 1. In the figure, \(x\), \(y\), and \(z\) are 3 stocks, representing the time integration of 3 rate variables, namely, the change rate of government promotion, the change rate of hospital participation, and the change rate of patient supervision, expressed by their respective replication dynamic equations; \(E_{g1}\), \(E_{g2}\), \(E_{h1}\), \(E_{h2}\), \(E_{p1}\), and \(E_{p2}\) are 6 intermediate variables, which respectively represent the expected return of each entity under different strategic states. The remaining 12 parameters are external factors outside the system boundary that may affect stock changes. The arrow lines indicate the causal relationship between 2 elements. The interaction of cause and effect acts as a causal feedback loop.

The purpose of system dynamics simulation is to reveal the law of development and change of things. The model construction focuses on the consistency, adaptability, and effectiveness of its overall structure. Therefore, the design of a system dynamics model does not require the authenticity and accuracy of parameter setting [19]. In this paper, common assignment methods are used for reference in parameter setting, and finally determined according to the experience of the interviewed experts. The specific parameter settings are as follows: initial time=0 years, final time=20 years, time step=0.125 years; \(R_p=30\); \(C_g=6\); \(K_g=3\); \(F_g=4\); \(R_h=50\); \(C_p=7\); \(K_p=5\); \(F_h=8\); \(M_h=10\); \(C_p=1\); \(F_p=2\); and \(R_p=6\). It should be noted that different initial values of each parameter will not change the simulation conclusion [20].

Figure 1. The system dynamics simulation model of the 3-party evolutionary game.
Ethical Considerations
Because no specific people or animals are involved in the research and no individual privacy is involved, academic ethics review is not required.

Results

Holistic Simulation Analysis of the Evolutionary Game Model

The initial state of the government, the hospital, and the patient in the evolutionary game is pure strategy, that is, each party’s strategy choice is 0 or 1, but the 8 possible equilibrium states formed by this are unstable. As long as one party’s strategy is slightly adjusted, the other parties will learn to imitate and then choose strategies with higher returns, and the whole game system will evolve into a new equilibrium state. Taking the (0, 0, 0) combination strategy as an example, the evolutionary game process under different values of $x, y,$ and $z$ is shown in Figure 2 (the ordinates in the figure represent the probability of strategy selection of the government, hospitals, and patients in the corresponding state).

Figure 2. Process of the evolutionary game in the case of combined values of $x, y,$ and $z$. (A) $x=0.01, y=0.01, z=0.01$; (B) $x=0.45, y=0.35, z=0.4$; (C) $x=0.99, y=0.01, z=0.01$; (D) $x=0.8, y=0.01, z=0.01$; (E) $x=0.05, y=0.5, z=0.01$; (F) $x=0.99, y=0, z=0.01$.

Comparing Figure 2A-C, we can see that as long as the government’s strategy evolves in the direction of vigorous promotion, no matter how much the initial probability of active participation of hospitals and patients is, the strategy of active participation and supervision will be adopted eventually, that is, as long as $X$ evolves from 0 to 1, the whole system will develop into a stable state of (1, 1, 1), and it shows that the evolution speed of the government’s strategy is faster than that of the hospital’s strategy. The evolution speed of the hospital’s strategy is faster than that of the patient’s strategy. At the same time, when the government’s strategy changes from 0 to 0.01 or from 1 to 0.99, it will eventually evolve into the strategy of trying to promote, showing that the strategy of trying to promote is always the best strategy of the government. During the vertical integration of cross-agency EHRs in the medical consortium, the government has always been and should be in the leading position. In specific practice, the government should take the initiative to promote rather than completely transfer the responsibility to the hospital.

When the government chooses to promote the strategy, the hospital will always reach the status of 1 before the patients. However, comparing Figure 2D,E, it can be found that when
the government chooses the negative treatment strategy, even if the hospital has high enthusiasm to participate in the vertical integration of EHRs at the beginning, it will gradually choose the negative participation strategy as time goes on; however, when the hospital completely chooses the passive participation strategy, even if the government adopts the promotion strategy again, it will be difficult to promote the hospital from the passive strategy to the active participation in a short time. The enlightenment from this conclusion is that the local government should do a good job in the overall design and development planning at the beginning, or it will be difficult to achieve the national goal of vertical integration of EHRs.

It can be seen in Figure 2 that the patients were the slowest to choose the active supervision strategy, but this is always the best strategy for patients. At the same time, it was also found that once patients choose the active supervision strategy, no matter how the strategies of the government and hospital change, the patients will tend to the active supervision strategy. Through the overall simulation analysis, it can be seen that the evolutionary game system will finally reach the equilibrium state at (1, 1, 1) after the 3 parties’ strategy mutation, learning imitation, and strategy adjustment in the game process.

Factors Influencing the Government’s Strategy Selection

According to the model simulation, 7 external variables such as the revenue $R_g$, cost $C_g$ when the local government tries to promote, the cost $K_g$ and punishment $F_g$ when the local government treats the hospital negatively, the government subsidy $F_h$ to the hospital, the supervision reward $F_p$ to the patients, and the punishment $M_h$ when the local government participates in the hospital negatively will affect the local government’s strategy choice (Figure 3). It can be seen from Figure 3A-C that $R_g$, $F_g$, and $K_g$ have similar effects on the government’s strategy selection. In the initial value state, increasing the values of the 3 parameters will make the government’s promotion probability $x$ reach 1 earlier, which means increasing the revenue when the government is trying to promote. By contrast, increasing the cost and punishment of the government’s negative treatment of the integration of EHRs will enhance the government’s enthusiasm for trying to promote. Comparatively, income has a greater impact on the choice of the government’s strategy. Comparing Figure 3D,E, it can be seen that $C_g$ and $M_h$ have similar effects on the government’s strategy. Increasing the values of the 2 parameters respectively will prolong the period for the government to choose the strategy of trying to promote, which indicates that the higher the promotion cost, the lower the probability of the government trying to promote. It is worth mentioning that when the punishment $M_h$ for passive participation in the hospital is increased to a certain extent, the government will slow down its efforts to promote the stability of the strategy, which indicates that the government cannot blindly increase its revenue by increasing the punishment. It also means that the government revenue $R_g$ pays more attention to the social value after the integration of EHRs rather than punishing the hospital. The change of $F_h$ and $F_p$ has no obvious impact on the government’s strategy. The curve trend after the change of the 2 parameters is basically the same, as shown in Figure 3E, which means that as long as the goal of integrating medical resources to improve medical efficiency can be achieved, the local government is likely to be willing to grant subsidies and supervise incentives. In other words, whether to give subsidies to hospitals or reward patients for supervision is not the main factor affecting the government’s choice of strategies.
Factors Influencing the Hospital’s Strategy Selection

Through simulation, it can be found that 4 external variables, namely, the cost $C_h$ when the hospital actively participates, the cost $K_h$ when the hospital passively participates, the government subsidy $F_h$ when the hospital actively participates, and the punishment $M_h$ when the hospital passively participates, will affect the choice of the enterprise strategy (Figure 4). When the value of $C_h$ is smaller or the values of $K_h$, $F_h$, and $M_h$ are larger, the probability of hospitals choosing the active participation strategy will be greatly increased. According to the analysis in Figure 4D, when the punishment $M_h$ for the government’s passive participation in the hospital increases to a certain extent, the marginal utility of the punishment measures will be greatly weakened, that is, when the punishment is increased to a certain extent, the incentive effect of the punishment itself will be weakened. A consistent conclusion has been reached in the analysis of the influencing factors of the government’s strategy choice.

It was also found in the simulation that the hospital’s revenue $R_h$ under normal conditions will not affect its strategy choice (Figure 4E). This shows that neither large hospitals that occupy an absolute dominant position in medical information nor small hospitals with limited resources will refuse the vertical integration of EHRs because of the amount of their own information resources. In other words, both large hospitals and small hospitals have the potential willingness to act collectively to achieve the vertical integration of EHRs. This finding is of great significance for the government to vigorously promote the vertical integration of EHRs (eg, electronic medical records).
Factors Influencing the Patient’s Strategy Selection

The factors that affect the patient’s strategy selection mainly include the cost $C_p$ and the reward $F_p$ when the patient actively participates in the supervision, as well as the social damage $R_p$ caused by the hospital’s negative participation (Figure 5). We can see in Figure 5A,B that when the patient’s supervision cost $C_p$ becomes larger or the government reward $F_p$ becomes smaller, the probability of patients choosing the active supervision strategy will obviously decrease. It can be seen from the simulation that the probability of patients choosing the active supervision strategy is determined by the difference between $F_p$ and $C_p$, as shown in Figure 5C. When $F_p>C_p$, the probability of patients choosing active supervision will significantly decrease. At this time, an increasing number of patients will take the “free riding” behavior and give up supervision, but when $F_p<C_p$, the probability of patients choosing active supervision is very low, and they will finally give up supervision. Therefore, to achieve the goal of vertical integration of EHRs, the government should provide some incentives for patient supervision and try to reduce the cost of patient supervision. When the social damage $R_p$ caused by the hospital’s negative participation is greater, the patients are more inclined to adopt the active supervision strategy, but this is not what we expect (Figure 5D). Therefore, the government is required to firmly choose and strive to promote the strategy to guide and encourage hospitals to choose the active participation strategy, so as to reduce the overall social damage.
Discussion

Principal Findings

Using evolutionary game analysis and system dynamics simulation methods, we analyzed the evolution process of long-term strategies of the government, hospitals, and patients in the vertical integration of EHRs, and the influencing factors and mechanisms of each party’s strategy choice. The principal findings are described below.

The choice of strategies of the 3 parties is jointly determined by a variety of external factors. Increasing the benefits of the government’s efforts to promote and increasing the costs and penalties of the government’s negative treatment have a greater impact on the government’s strategic choice, while subsidies to hospitals and incentives to patients only have a little impact on the government’s strategic choice. The following scenarios would motivate hospitals to choose an active participation strategy: lower cost of active participation, higher cost of passive participation, higher government subsidies, and reasonable penalties. The revenue of hospitals under normal circumstances have a little impact on the choice of hospital’s strategies. Whether patients choose the active supervision strategy is mainly determined by the difference between the government reward and the supervision cost. Only when the patient supervision reward is higher than its cost can the purpose of encouraging patients to actively supervise be achieved.

The vertical integration of EHRs within the medical consortium has practical feasibility. In the model simulation, it is found that the strategies of the government, hospitals, and patients finally reach the optimal equilibrium state at (1, 1, 1), and both large and small hospitals have the potential willingness to take unified action to achieve the vertical integration of EHRs. However, this requires the government to be in a leading position and make every effort to promote it. It is difficult to achieve the goal of vertical integration of EHRs if the integration responsibility is fully delegated to the hospital.

Constructing a multiagent coordination mechanism under the guidance of the government is the only way to achieve the goal of vertical integration of EHRs in the medical consortium. The research shows that the government’s strategy is influenced by strategies of both hospitals and patients. Patient’s strategy is also influenced by the strategies of both the government and the hospital; the hospital’s strategy choice is mostly directly affected by the government’s strategy, and the patient’s strategy can only indirectly affect the hospital’s strategy choice through the impact on the government’s strategy. Establishing a good cooperation mechanism can not only improve the cooperation efficiency, but also reduce the participation cost, so as to accelerate the realization of the integration goal. Therefore, it is necessary to build a multibody coordination mechanism led by the government and fully mobilize the enthusiasm of all parties so that the medical consortium can achieve the goal of vertical integration of EHRs.

The simulation results revealed that the appropriate increase of government subsidies for active participation and penalties for passive participation can enhance the enthusiasm of hospitals, while the improvement of supervision incentives can also encourage patients to choose active supervision strategies. Therefore, the government should establish a reasonable reward and punishment mechanism, and design a detailed implementation plan. The government should also build a

Figure 5. Impact of external variables on the patient’s strategy selection. (A) Effect of $C_p$ on the patient’s strategy selection; (B) Effect of $F_p$ on the patient’s strategy selection; (C) Joint effect of $F_p$ and $C_p$ on the patient’s strategy. (D) Impact of $R_p$ on the patient’s strategy selection.
reasonable benefit distribution mechanism among medical institutions to ensure the mutual use performance after the vertical integration of EHRs. Performance evaluation plays an important role in guiding, supervising, and promoting the vertical integration, and is also the main basis for establishing a reward and punishment mechanism and a benefit distribution mechanism. Therefore, the government should independently design the performance evaluation mechanism of information resource integration in the evaluation system of medical consortium construction, to reduce “free riding” opportunistic behavior and improve the enthusiasm of hospitals and patients to participate.

Limitations
Admittedly, there are limitations in this study. First, only the 3 core participants (the government, hospitals, and patients) were considered, while other stakeholders (eg, technology providers) were not considered. Second, because of the constraints of research conditions, real data available for prediction and simulation were not obtained. In the future, we will consider a wider range of stakeholders and seek to obtain real data for empirical studies to further improve the reliability of our findings.

Conclusions
We should build a multiagent coordination mechanism. As discussed in the preceding sections, strategies of the government, hospitals, and patients are intertwined and each is influenced by multiple external factors. Achieving the goal of vertical integration of EHRs requires the participation of multiple parties, with the government in a leading position. Establishing a good coordination mechanism can improve the efficiency of collaboration and reduce the cost of participation, thus accelerating the achievement of integration goals. Therefore, it is imperative to establish a government-led multiparticipant collaborative mechanism.

We should set up a reward and punishment mechanism and a benefit distribution mechanism. Simulation results show that appropriately increasing government subsidies when hospitals are actively involved and penalties when they are negatively involved can increase hospital motivation. Increasing monitoring rewards can also motivate patients to choose active monitoring strategies. Therefore, the government should establish a reasonable reward and punishment mechanism and design a detailed reward and punishment implementation plan. It should also build a reasonable benefit distribution mechanism among medical institutions to ensure the interoperability performance of vertically integrated EHRs.

We should establish a scientifically integrated performance evaluation mechanism. Performance evaluation aims to compare, measure, and judge the gap between the real development status and the predefined expectation standard. The evaluation results have an important role in guiding, supervising, and promoting the development of EHRs toward verticality, and are also the main basis for setting up reward and punishment mechanisms and benefit distribution mechanisms. Therefore, the government should independently design the performance evaluation mechanism of information resource integration in the evaluation system of medical consortium construction, so as to reduce the opportunistic behavior of “free riding” and increase the enthusiasm of hospital participation.

Acknowledgments
This work was funded by the National Natural Science Foundation of China (grant numbers 72164037 and 71864021).

Authors’ Contributions
SHT contributed to model construction and simulation analysis, and wrote the original draft. YC reviewed and edited the manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

**EHR:** electronic health record
Abstract

Background: The COVID-19 pandemic introduced an urgent need for effective strategies to disseminate crucial knowledge and improve people’s subjective well-being. Complementing more conventional approaches to knowledge dissemination, game-based interventions were developed to create awareness and educate people about the pandemic, hoping to change their attitudes and behavior.

Objective: This study provided an overview and analysis of digital and analog game-based interventions in the context of the COVID-19 pandemic. As major pandemics and other large-scale disruptive events are expected to increase in frequency in the coming decades, this analysis aimed to inform the design, uptake, and effects of similar future interventions.

Methods: From November 2021 to April 2022, Scopus, Google, and YouTube were searched for articles and videos describing COVID-19–themed game-based interventions. Information regarding authorship, year of development or launch, country of origin, license, deployment, genre or type, target audience, player interaction, in-game goal, and intended transfer effects was extracted. Information regarding intervention effectiveness was retrieved where possible.

Results: A diverse assortment of 23 analog and 43 digital serious games was identified, approximately one-third of them (25/66, 38%) through scientific articles. Most of these games were developed by research institutions in 2020 (13/66, 20%) and originated in Europe and North America (38/66, 58%). A limited number (20/66, 30%) were tested on relatively small samples, using a diversity of research methods to assess the potential changes in participants’ knowledge, attitudes, and behaviors as well as their gameplay experience. Although most of the evaluated games (11/20, 55%) effectively engaged and motivated the players, increased awareness, and improved their understanding of COVID-19–related issues, the games’ success in influencing people’s behavior was often unclear or limited.

Conclusions: To increase the impact of similar future interventions aimed at disseminating knowledge and influencing people’s attitudes and behaviors during a large-scale crisis, some considerations are suggested. On the basis of the study results and informed by existing game theories, recommendations are made in relation to game development, deployment, and distribution; game users, design, and use; game design terminology; and effectiveness testing for serious games.

(JMIR Serious Games 2023;11:e41766) doi:10.2196/41766

KEYWORDS
COVID-19; serious game; game-based intervention
**Introduction**

**Background**

The outbreak of the novel coronavirus causing COVID-19 was declared a pandemic in March 2020 [1]. With increasing restrictions and uncertainty, a growing number of studies have reported troubling large-scale psychological effects of the COVID-19 pandemic, including worry, fear, anxiety, boredom, frustration, irritability, anger, sadness, and depression [2-7]. On a collective level, health anxiety, collective disorientation, community panic, crisis fatigue, and social isolation fatigue have been observed [3,4,8]. These effects seem to be perpetuated by doomsurfing or doomscrolling, which refers to the tendency to continually search for bad news on the web even though it is saddening, disheartening, or depressing [9]. An increase in additional dysfunctional behaviors such as panic purchasing of essential household items, increased consumption of psychoactive substances, weight gain because of stress-induced eating, and complete social withdrawal has suggested an imbalance between increased environmental demands and the efficacy of people’s resilience and coping. Experts have advised that people adhere to daily routines and maintain a strong degree of social connectedness to improve communication and reduce boredom [10-12]. Less conventional means to maintain or improve people’s subjective well-being have also been investigated, including digital health interventions (ie, telemedicine, eHealth, and mobile health) [13-15], exergames [16], arts-based interventions [10], bibliotherapy [17], and the company of pet animals [18].

Playing games has been a popular coping strategy during the pandemic [19,20]; it has even been recommended by the World Health Organization (WHO) as an effective way to stop the spread of COVID-19 [21]. Kleinman et al [22] reported that people resorted to games to maintain their mental well-being (eg, to cope with loneliness, escape from the troubling real world, or replace the loss of routines because of quarantine), connect with others (eg, with distanced friends and family, people confined to the same household, or new people), or substitute reality (eg, using game characters as a substitute for real interactions and hosting internet-based events). Some people have also found solace in creating games or gamifying everyday activities to cope with distress during quarantine [22].

During the pandemic, an increased interest specifically in pandemic-themed games such as the board game Pandemic, which was developed after the 2003 severe acute respiratory syndrome epidemic, was observed [23]. In 2020, over 50 global game industry leaders launched the campaign #PlayApartTogether to encourage gamers to follow the WHO guidelines as well as incorporate COVID-19 prevention messages into their games [24,25]. Thus, in addition to entertainment games intended to be played primarily for amusement purposes, a considerable number of serious games were developed to educate people about the COVID-19 pandemic, influence their behavior, and improve their well-being in terms of increased life satisfaction and decreased negative affect. These games drew upon the proven positive effect that (serious) games can have on increasing health literacy [26,27].

Although the entertainment in playing games can in itself benefit one’s psychological well-being, our study focuses on serious games that were intentionally designed to spread awareness and knowledge of COVID-19, promote healthy behaviors, and increase people’s resilience during the pandemic.

**Objectives**

Recent reports such as that by Metabiota [43] have warned that the risk of another major pandemic during our lifetimes is higher than many expect—they estimate the likelihood of another pandemic occurring within the next 25 years to be 47%-57%. Similarly, Marani et al [44] predicted up to a 3-fold increase in the yearly probability of extreme epidemics in the coming decades. This implies that the games developed in response to the recent pandemic could potentially serve as a basis for future games, indicating a need to explore the landscape of COVID-19–themed games. Apart from a recent analysis of 5 studies that evaluated 4 games in total [45], to date, no comprehensive reviews that could help evaluate the various game design decisions for this specific context have been published. Thus, the overview presented in this paper aims to inform the design of similar game-based interventions by showcasing the diversity of intentions and design strategies.
Methods

Search and Selection of COVID-19 Games
In November 2021, we performed a search of Scopus to identify scientific articles that described the development or testing of COVID-19–themed games. The search was limited to articles and conference papers published in English after 2018 that included the keyword “game*” and any of the keywords “covid*,” “sars-cov*,” “pandemic*,” or “lockdown*” in the title. To further reduce the number of irrelevant results, we excluded those where the word game was used in relation to game changer or the Olympic and Paralympic games. The search was then repeated in April 2022 to include the most recent studies. Titles and abstracts of the identified 181 results were screened to exclude publications that did not focus on games developed to educate the general public on the COVID-19 pandemic. In total, 34 full texts describing 25 different games were included in the review.

An important consideration that informed the literature review was that games are not necessarily disseminated through or discussed in scientific publications. To avoid excluding relevant games that had only been disseminated through nonscientific media, Google and YouTube were also searched using the aforementioned keywords. This search, which was performed from November 2021 to April 2022, identified 42 additional games developed by private citizens, studios, and profit and nonprofit organizations. Owing to the overwhelming number of prototypes of digital games developed primarily for game jams (eg, International Festival of Independent Games “JAMMING THE CURVE: COVID-19 Game Jam”) or competitions (eg, the 2021 Institute of Electrical and Electronics Engineers Conference on Virtual Reality and 3D User Interfaces contest “Challenging Pandemics”), these were not included in this study.

Data Extraction
The identified games were analyzed and categorized based on the Comprehensive Taxonomy for Serious Games by De Lope and Medina-Medina [46]. This taxonomy builds on a series of concepts from existing partial classification systems to collect and organize a large number of game features when developing new games or when analyzing or comparing existing ones [46]. Information regarding the games’ authorship, license, deployment, genre or type, target audience, player interaction, in-game goal, and intended transfer effects was extracted. Information regarding the year of development or launch, country of origin, and user testing was also obtained where possible.

Results

COVID-19 Games

Overview
By January 2022, a total of 66 COVID-19–themed games—23 (35%) analog and 43 (65%) digital—were identified. Detailed overviews of analog and digital games are provided in Multimedia Appendix 1 [47-52] and Multimedia Appendix 2 [53-70], respectively; Figures 1-10 show 10 game examples. The general findings are presented in the following sections.

Figure 1. Clinic Deluxe Edition: CoVid_19 variant (image reproduced with permission from Alban Viard [71]).
**Figure 2.** Plague Inc.: The Cure (image reproduced with permission from Ndemic Creations [72]).

**Figure 3.** Antidote COVID-19 (image reproduced with permission from Psyon Games [73]).
**Figure 4.** At-Home Scavenger Hunt [47]. Materials developed by the Centers for Disease Control and Prevention (CDC). Reference to specific commercial products, manufacturers, companies, or trademarks does not constitute its endorsement or recommendation by the US Government, Department of Health and Human Services, or Centers for Disease Control and Prevention.

**At-Home Scavenger Hunt**

**INSTRUCTION SHEET — to be used by facilitator**

**Instructions:** Encourage items to be searched for at home. You don’t need to go outside, which is a big deal! Do not forget to wash each item thoroughly. The first person to bring the item to the host seated for the item, which is stored in the following order:

- Headache
- Flu
- Cough
- Diarrhea
- Fever
- Sore throat
- Redness/irritation of eyes
- Pain/pressure in the chest
- Shortness of breath
- New loss of taste or smell
- Chills

- **Neurological symptoms**
  - Confusion
  - Blurred vision
  - Weakness
  - Loss of coordination
  - Numbness

- **Blooduffy multisystem inflammation**
  - Abdominal pain
  - Muscle weakness
  - Diarrhea
  - Vomiting

- **Physical activity**
  - Low-grade fever
  - Dizziness
  - Nausea
  - Loss of appetite
  - Cough
  - Difficulty breathing

- **Factors related to products**
  - Paper towels
  - Sanitizer
  - Food
  - Trash can
  - Laundry


**Figure 5.** GO VIRAL! (image reproduced with permission from Tilt Studio and Professor Sander L van der Linden from the University of Cambridge [66]).
Figure 6. You Make Me Sick! (image reproduced with permission from The Partnership in Education [52]). The game was created by The Partnership in Education, a project supported by the National Institute of General Medical Sciences of the National Institutes of Health under award R25GM132910.
Figure 7. Lockdown! (image reproduced with permission from Yann Boucher [76]).

Figure 8. Corona Bee (image reproduced with permission from Focus Games [56]).
Figure 9. Heroes of Covid-19 (image reproduced with permission from GRM Digital [67]).

Figure 10. Destroy COVID (image reproduced with permission from Kristóf Horváth [74]).
Release Year and Location
The vast majority of the identified games were developed or launched in 2020 (41/66, 62%); the rest were developed or launched in 2021. The largest number of games originated in the United States (10/66, 15%), India (9/66, 14%), and the United Kingdom (7/66, 11%). The countries of origin of other games included Denmark, the Netherlands, and Switzerland (3/66, 5% of the games each); Canada, Czech Republic, Finland, Germany, Morocco, Portugal, and South Africa (2/66, 3% of the games each); and Australia, Bangladesh, Belgium, Brazil, China, France, Hungary, Indonesia, Poland, Saudi Arabia, Singapore, and Taiwan (1/66, 2% of the games each). For 8% (5/66) of the games, the country of origin could not be determined.

Authorship
The lead developers were research institutions (31/66, 47% of the games), game development studios (17/66, 26% of the games), or private individuals (13/66, 20% of the games). A total of 9% (6/66) of the games were developed by children aged 10 to 17 years (Better than Hugo, Corona – Mit Eifer ins Geschäft, Corona Yuga, COVID-19: A Race to the Vaccine, Go Corona Go, and Infected!). Governmental and nongovernmental organizations (eg, ministries, educational institutes, the WHO, the Centers for Disease Control and Prevention [CDC], the United Nations Children’s Fund, and Médecins Sans Frontières) were involved in the development of 22% (5/23) of the analog games and 16% (7/43) of the digital games.

License
In total, 65% (15/23) of the analog games were commercial (price: US $6.54-33.77), 30% (7/23) were available for free (6/7, 86% of these as Print & Play games), and 9% (2/23) were Kickstarter prototypes. In contrast, the vast majority of digital games were free to play (20/43, 47%) or prototypes (20/43, 47%); 5% (2/43) were commercial (price: US $4.36-9.80), and 2% (1/43) were developed for internal use at a company.

Target Audience
Most games (38/66, 58%) targeted younger audiences, including children as young as 3 years old, teenagers, and university students. Analog games were often designed to be engaging for the entire family, whereas 36% (24/66) of all digital games did not have a defined target audience. Certain other games targeted very specific user groups, such as gamers, active social media users, health care workers, health enthusiasts, employees of certain enterprises, and the African population.

Game Types and Genres
Practically all games identified were simulations of the COVID-19 pandemic. Some were set in the players’ immediate environment, such as the classroom (eg, Fighting COVID-19 at Purdue University), the office (eg, Social Distancing – The Game), or the supermarket (eg, Dino-Store), or more broadly in a village, town, or city (eg, Clinic Deluxe Edition: CoVid_19 variant [Figure 1], COVID Dodge, SurviveCovid-19, and VRS Fight Club); a country (eg, Better than Hugo and Korona Ira); or the world (eg, Plague Inc.: The Cure; Figure 2), where the players were tasked with controlling the spread of the disease. Others were set in a microscopic environment of the human body, where the players assumed the role of immune cells in fighting off the infection (eg, Antidote COVID-19 [Figure 3] and Infekcja) or of the virus in spreading the disease (eg, Viruscape). Other interesting approaches were used in games such as At-Home Scavenger Hunt by the CDC (Figure 4), which required the players to find relevant physical items in their actual environment; the location-based game MeetDarius, where virtual items could be collected in the player’s physical environment (similar to Pokémon GO) if the player was wearing a face mask, the presence of which was detected by face recognition software; and GO VIRAL! (Figure 5), which was set in the environment of a simulated social media platform.

The vast majority of analog games were board games (15/23, 65%) or card games (6/23, 26%). The most common analog game types or genres were strategy games (11/23, 48%; eg, Lockdown!; Figure 7) and race-to-the-end games (6/23, 26%; eg, You Make Me Sick!; Figure 6), whereas most digital games were primarily strategy (15/43, 35%; eg, Plague Inc.; The Cure), action (14/43, 33%; eg, COVID Dodge), and trivia (12/43, 28%; eg, Corona Bee; Figure 8) games. Some less common game types such as meditation or Zen (CoronaShield), hidden object (Help to stop the COVID-19 coronavirus and At-Home Scavenger Hunt by the CDC), memory (Heroes of Covid-19; Figure 9), and escape room–type puzzles (COVID-19 [CORONA VIRUS] and Destroy COVID; Figure 10) were also identified.

Deployment
Most digital games (23/43, 53%) were web-based and playable on PCs, mobile devices, consoles, or a combination of these. An additional 21% (9/43) were deployed for mobile platforms, 7% (3/43) were deployed for desktop and mobile platforms, 2% (1/43) were deployed for desktop only, 2% (1/43) were deployed for desktop and consoles, and 2% (1/43) were deployed for consoles only. Analog games mainly took the form of physical objects (eg, game boards with figures and dice, play money or tokens, playing cards, and puzzle pieces) or free electronic files that were designed to be printed at home (Print & Play).

Player Interaction
Of the 43 digital games, 37 (86%) were single-player, 4 (9%) were multiplayer (up to 7 players), and 2 (5%) were single- and multiplayer. Of the 6 multiplayer digital games, 5 (83%) were co-operative, and 1 (17%) was competitive. In contrast, all analog games (23/23, 100%) were multiplayer (up to 100 players), although single-player versions were also available in 26% (6/23) of the cases. Of these 23 games, 18 (78%) were competitive, 4 (17%) were co-operative, and 1 (4%) involved only teams or partnerships.

Gameplay Duration
Information retrieved from 15 analog games showed that gameplay duration ranged from 5 to 180 minutes—3 (20%) games required up to 15 minutes, 5 (33%) games required up to 30 minutes, 2 (13%) games required up to 45 minutes, 2 (13%) games required up to 60 minutes, and 3 (20%) games required >1 hour. The estimated duration could only be retrieved for 7% (3/43) of the digital games, which were substantially shorter than most analog games and lasted 5 to 15 minutes.
Intended Transfer Effects

Overview
In addition to a mostly entertaining in-game goal, serious games explicitly add a more serious, out-game goal [75], which refers to the intended effect of the game in the real world [28]; such an effect can be observed after playing the game and may include increased knowledge and awareness, skill development, and changes in attitudes or behaviors. Such serious games aim for impact that is typically long-term, in contrast to more short-term or direct forms of impact such as interactions, actions, and experiences while playing the game (for discussion on human impact–centered design, refer to the study by Fokkinga et al [77]). This repertoire of COVID-19 games covers this diversity of purposes, as summarized in Table 1 (extracted from the games’ descriptions; for details, refer to Multimedia Appendix 3).

Table 1. Intended transfer effects of COVID-19–themed games (listed according to the hierarchy of effects model originally developed for advertising and recommended by McGuire [78] for use in public health campaigns [79]; N=66).

<table>
<thead>
<tr>
<th>Intended transfer effect</th>
<th>Games, n (%)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Awareness-oriented effects</strong></td>
<td>11 (17)</td>
<td>- Antidote COVID-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Beat Corona</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Corona – Mit Eifer ins Geschäft</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- COVID Challenge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Heroes of Covid-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Social Distancing – The Game</td>
</tr>
<tr>
<td><strong>Education-oriented effects</strong></td>
<td>38 (58)</td>
<td>- Antidote COVID-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Clinic Deluxe Edition: CoVid_19 variant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Corona Bee</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- COVID Safety Simulation: CAMPUS LIFE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Govid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- You Make Me Sick!</td>
</tr>
<tr>
<td><strong>Attitude-oriented effects</strong></td>
<td>12 (18)</td>
<td>- Better than Hugo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Breaking the Magic Circle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Corona Game</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Essential Workers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- GO VIRAL!</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SurviveCovid-19++</td>
</tr>
<tr>
<td><strong>Behavior-oriented effects</strong></td>
<td>8 (12)</td>
<td>- At-Home Scavenger Hunt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Corona Yuga</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- CovidShield Game Suite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Fighting COVID-19 at Purdue University</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- MeetDurian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- The Magic Soldier of the Human Body</td>
</tr>
</tbody>
</table>

Increasing Awareness and Spreading Knowledge
Overall, the games reviewed in this study aim to educate the players on the COVID-19 pandemic in an entertaining way. Practically all games (66/66, 100%) were developed to spread general awareness about COVID-19 (eg, Is COVID-19 caused by bacteria or a virus?; Name one symptom of COVID-19 that affects the respiratory system; and What does the “19” in COVID-19 stand for?). Some (6/66, 9%) were intended to educate people about the means and mechanisms of virus spread, the immune response, or both (eg, True or False: COVID-19 can only be spread from person-to-person and Can you get COVID-19 from drinking water?), whereas others (34/66, 52%) were intended to educate people about the means and mechanisms of virus spread, the immune response, or both (eg, True or False: COVID-19 can only be spread from person-to-person and Can you get COVID-19 from drinking water?), whereas others (34/66, 52%) focus more on safety precautions to avoid catching and spreading the infection (eg, True or False: Wearing a face mask can help stop the spread of COVID-19 to others; When should you wash your hands?; and What is the minimum percentage of alcohol that a hand sanitizer must have?). Still others (5/66, 8%) specifically address vaccination (eg, True or False: The COVID-19 vaccines can give you COVID-19; True or False: The vaccine will damage my DNA; I’m healthy and I’m never ill, should I still get vaccinated?; and Should a pregnant woman have the vaccine?).

Influencing Attitudes
Several games (11/66, 17%) aim to cultivate people’s sense of empathy and collective responsibility (eg, Clinic Deluxe Edition: CoVid_19 variant, Fighting COVID-19 at Purdue University, and Essential Workers) as well as encourage critical reflection on policy makers’ dilemmas (eg, Better than Hugo and...
Lockdown! or the spread of misinformation (eg, CoronaChampion and GO VIRAL!). Heroes of Covid-19 may be the only game to focus on positive aspects of the pandemic, raising awareness of people’s noble deeds during this period (eg, “Quinn Callendar, a boy scout from Canada, used his 3D printing skills to create ear guards for people who feel pain from wearing masks all day”; “Intensive care nurse Molly Watts created a book entitled ‘Dave the Dog is worried about coronavirus’ to help tackle anxiety in children amid the Covid-19 outbreak”; and “Annemarie Plas from Brixton, London united the nation and gave the NHS frontline a boost with her ‘Clap for our Carers’ initiative.”).

**Influencing Behavior**

Some games were designed to promote adherence to safety precautions in real-life settings. For example, Fighting COVID-19 at Purdue University aims to encourage students to clean their university laboratories, and MeetDurian encourages players to wear protective masks in public. Other games aim to stimulate behavior that can directly improve the players’ well-being, such as mindfulness breathing (CovidShield Game Suite) and immunity-increasing nutrition (The Magic Soldier of the Human Body).

In most cases, the players were to adopt the positive role of a hero fighting the virus or protecting their avatar from infection in a simulated environment. An interesting yet less common alternative was to teach people about undesirable or unsafe behaviors by putting them in the negative role of, for example, the virus (Virascape), an infected customer at a store (Instructional remote multiplayer VR game), or a misinformation spreader (GO VIRAL!).

**Testing of Effectiveness in Promoting Transfer Effects**

A relatively small number of scientific studies that investigated the effectiveness of the individual COVID-19–themed games (20) have been published (Multimedia Appendix 4 [80-101]). In total, 3 categories of effects were most frequently tested: the change in participants’ knowledge of COVID-19 and infection prevention measures, their gameplay experience, and the games’ potential to facilitate players’ attitude or behavior change. The studies used various research methods and tools, including heuristic evaluations by experts, surveys composed of custom or standardized questionnaires and administered before or after gameplay, observations during gameplay, semistructured interviews, self-reports, or game log analysis after gameplay. A total of 15% (3/20) of the studies assessed the players’ changes in knowledge [80-82] and attitudes [80,83,84] by comparing their pre- and postgameplay responses. One study involved a control group that watched an educational video about COVID-19 for comparison with a test group that played the game [80], whereas another used comparative analysis with other similar games [85]. With the exception of 30% (6/20) of the studies, which involved larger numbers of participants or instances of gameplay, testing was performed on relatively small samples (2-30 participants).

At this point, it is important to emphasize the diversity of methods used to assess the effectiveness of the games as these can greatly limit the comparability of the results. Nevertheless, most of the studies considered (11/20, 55%) found that the games succeeded in engaging and motivating the players [80,81,83,84,87,91,94,99,101], increasing awareness, and facilitating teaching COVID-19 hygienic knowledge [80-83,89,91]. Improved understanding led to positive changes in players’ attitudes toward COVID-19 preventive measures [80,82,83,87,89,91,94,97,102]. A study reported that health anxiety remained relatively unchanged in response to playing 1 game [80].

Certain limitations of these studies have been acknowledged, especially in relation to the games’ capacity to change players’ behaviors. A proper assessment of the games’ capacity to influence behavior was found to be particularly challenging as different sources of information may have contributed to people’s adherence to safety measures [90,103] and as other factors—most notably, the players’ ages—also affected the games’ effectiveness [87,94]. The effectiveness of the reviewed games was often unclear or limited [98,103], and Suppan et al [92] explicitly stated that a longitudinal trial would be necessary to accurately assess their behavioral impact.

In addition, the studies recruited participants to play the games and evaluate their impact as opposed to surveying people who had already obtained and played the games through their own initiative. Information regarding the number of times the games were downloaded or played also does not appear to be structurally documented or made available; thus, it remains unclear to what extent the games were able to motivate people to play them in the first place. However, this information is of key importance as a game cannot be effective if it is not played.

**Discussion**

**Principal Findings**

This study aimed to provide an overview of games that were developed during the COVID-19 pandemic to disseminate crucial knowledge and enhance people’s subjective well-being. In general, we observed an extensive proliferation of games—especially digital games—in 2020, when the pandemic was declared. Most (31/66, 47%) were developed by research institutions in Europe and North America—some in collaboration with health authorities. All efforts notwithstanding, the overall frequency of use and impact of these games seem to have been modest at most.

To effectively influence people’s attitudes and behavior, design interventions need to be tailored to each situation, considering the environmental demands—as perceived by the target group—and their personal resources to meet these demands. On the basis of the findings of this study, we provide recommendations for the design of games aimed at improving people’s well-being in future global health crises similar to the COVID-19 pandemic.

**Game Development, Deployment, and Distribution**

**Overview**

A game only has an impact if it is played. It is noteworthy that we only became aware of the multitude of COVID-19–themed games when we started actively searching for them. Although
our experience is anecdotal, it does suggest the importance of investing time and effort in the dissemination of serious games.

**Develop a Dissemination Plan**

In the context of a pandemic, game developers should carefully consider their target audience. In many cases, this is probably the broadest audience possible, making it crucial to plan for the effective dissemination of the game. Most of the reviewed analog games (15/23, 65%) were commercially available physical objects, whereas most digital games (20/43, 47%) were freely accessible web-based applications. Each of the various forms of dissemination has advantages and disadvantages, but deployment strategy and cost may influence game accessibility to the greatest extent. Commercial games, for example, have an advantage in that they have access to professional dissemination channels and are often designed to be intriguing and esthetically appealing to attract people’s attention. However, the cost and effort necessary to obtain such games can preclude them from reaching a broader audience. For analog games, *Print & Play* games tend to be the most convenient to access, especially when freely available on the web. Such games have a relatively low threshold for access in that buying them is not necessary.

When social distancing is required, as was the case during the pandemic, playing analog games is typically only possible for members of the same household. In contrast, digital games are easy to obtain and play remotely provided that the users have access to electricity and the internet. To maximize reach, Hill et al. [82] advised that COVID-19–themed games must be platform-independent and playable with minimal hardware. Collaborating with health authorities, the media, nongovernmental organizations, educational institutions, (local) governments, and social media influencers can also help spread awareness of a game’s existence, although this may not always be the case. The designers of *Escape COVID-19*, for example, conducted a retrospective analysis to investigate the reasons for their limited success in recruiting participants for their study [104]. Specifically, they aimed to assess the effect of 3 different dissemination strategies on game account creation over a period of 6 months. In the first period (53 days), the game was disseminated by a part-time worker; following this was a press release (15 days); and in the final stage, the game was officially announced by the Swiss Federal Office of Public Health (15 days). Their findings suggested that the press release was the most successful and the official communication was the least successful. Nonetheless, the sequence and duration of communication interventions should also be considered.

**Ensure the Reliability of Information**

For games that aim to inform and share recommendations during a pandemic, a prominent challenge is that such recommendations are inherently dynamic. Although some information might be fixed (eg, what a virus is), other information may likely change after a game is introduced. The use of reliable sources such as the WHO, the CDC, and local governments can assure people that the communicated knowledge is factual. Practically all the reviewed games based their learning objectives on such sources. However, as guidelines may differ among different institutions and game developers may have different opinions on which information is correct, it is recommended that players be explicitly informed about the sources of the communicated knowledge. In addition, the credibility of the aforementioned sources can be subject to change, as seen with the rapidly waning trust in the CDC in the United States.

Real-time updates can be especially helpful in tackling confusion when guidelines change rapidly owing to the evolving understanding of health crises. It is critical that pandemic games be designed in a way that prevents them from becoming outdated too quickly; this can be addressed by incorporating sufficient flexibility in game design to allow for changing guidelines or information. Although this tends to be more challenging for analog games, it is possible—the content of analog games can be updated via expansion packs, and hybrid games where physical game elements are complemented by a regularly updated mobile app can be developed.

**Game Users, Design, and Use**

How, where, when, and with whom a game is intended to be played can influence players’ motivation and the efficiency of the learning process [105]. Thus, it is important to design pandemic games that facilitate the acquisition and transfer of the intended learning objectives under pandemic-specific learning conditions.

**Consider the Target Audience and Context of Use**

**Overview**

When designing any serious game, it is generally advised to consider the entire ecosystem of stakeholders, including not only the players but also the initiators of games (eg, governments, nongovernmental institutions, health authorities, schools, and private citizens), design agencies, and game developers. The involvement of these stakeholders in various phases of game design (ie, problem definition, product design, and tailoring) can improve a game’s implementation [106]. For pandemic games, this can be challenging, especially if the game is intended to appeal to a broad variety of affected people. Therefore, it is necessary to determine exactly which stakeholders need to be targeted. The following sections focus specifically on player characteristics.

**Player Demographics**

According to the effectiveness studies we reviewed, certain target group characteristics—most notably, age—influence people’s engagement and the perceived effectiveness of the games [87,94]. Thus, caution is required when designing multiplayer games for people of different generations (eg, family games); it is important to avoid making such games too complex for very young or inexperienced audiences or too simple to effectively motivate adults and experienced players.

A general recommendation for game design is that game instructions should be easy to understand for a targeted user group. This recommendation is especially important for pandemic-focused games to minimize as much as possible the chances of spreading misinformation or confusion. Special attention needs to be paid to the language (preferably native language or multilingual versions) and vocabulary (preferably
simple and familiar expressions) in the games. Using rules or elements from popular existing games (eg, Monopoly, Ludo, Uno, and Pacman) can help users understand and adopt a game more rapidly as people can only appreciate novelty when the product is also simultaneously perceived as familiar or typical [77,107].

**Player Type**

The process of meaning attribution can be influenced by the players’ culture, personality traits, and current mental state, as well as by other people [77]. Similarly, product emotions, which are the emotions users experience in response to (playing) the game, are influenced by their personal needs, goals, values, and abilities [77]. Thus, a detailed psychological understanding of a game’s target population can increase its impact.

First, it is important to understand that different people choose to play games for different reasons. A popular typology of player personalities developed by Bartle [108] distinguishes between 4 player types that differ based on their preferred way of engaging in the virtual world: achievers, explorers, socializers, and killers (for details, refer to the study by Bartle [108]). People assume different styles of play depending on their mood or in-game goals [108], suggesting that setting appropriate goals and using game elements to influence players’ moods may discourage adverse social in-game behavior. However, Bartle based his taxonomy specifically on multiuser dungeon games; thus, his taxonomy cannot be directly extrapolated to other game types. The player profiles by Bartle [108] might work for entertainment games but not for serious games. To address this issue, Siriaraya et al [28] advised that users’ experiential preferences be investigated for the game world as well as for the real world.

For COVID-19 games, it is also important to consider the differences between social and solitary players, especially in the degree of autonomy, presence, and relatedness they seek to experience (for details, refer to the study by Vella et al [109]). Understanding the specific needs and sources of motivation for different players can support the preferred player interaction.

**Support Player Interaction**

All the analog games reviewed in this study (23/23, 100%) were multiplayer and mainly competitive, whereas only 14% (6/43) of the digital games supported a multiplayer mode, of which 83% (5/6) were co-operative. The main advantage of single-player games is that they can be played at any time independently of other people. However, most single-player games do not allow for socializing, sharing opinions, or discussing ideas. In contrast, multiplayer games allow for socializing, but the minimum number of players required can limit the number of opportunities to play the game as this may depend on other people’s interest, availability, or both. Optimally, a game should be designed to support both single- and multiplayer modes, as was the case for 26% (6/23) of the analog games and 5% (2/43) of the digital games studied in this review.

During the COVID-19 pandemic, collective action and compliance with safety guidelines were paramount in reducing the spread of the virus. Co-operative games can communicate this need to the players and support social bonding and collective decision-making that can be transferred from the in-game world to real-life situations. Previous studies have suggested that participation in co-operative—as opposed to competitive—multiplayer games can support team building [110-113] and increase real-world collaboration and prosocial behaviors among players [114-117]. In contrast, the motivation to win in competitive games may leave little room for bonding and the exchange of ideas, possibly limiting the encouragement of prosocial behavior. Nevertheless, competitive games tend to be more stimulating, whereas co-operative games may support passive participation when responsibilities are not equally distributed among the players. Therefore, it is advised to design for teams or partnerships, which combines both co-operation among team members and competition against the opposing team or teams.

**Balance Visual Design and the Realism of the Simulated Domain**

Immersive simulations of realistic environments and experiences where acquired knowledge can be put to use without the serious consequences of a real disaster may facilitate players’ transfer of new knowledge to everyday life [29]; this is particularly important during pandemics, when rapid changes in attitudes and behavior are required. Presumably for this very reason, practically all the reviewed games were either literal or metaphorical simulations of the COVID-19 pandemic. However, as such games tend to emphasize educational goals, caution is necessary to adequately balance such education with entertainment so that games remain appealing to a broader audience.

To foster players’ motivation and consolidate their engagement, fun elements that are unrelated to the simulated domain (eg, avatar shape or a game within a game) can be used. For example, the authors of Escape COVID-19 found that attractive graphics that were adapted to fit the preferences of the target population were essential in increasing players’ engagement [92]. Nevertheless, the designers of Point of Contact [82] stressed that a game’s graphics and cross-player communication need to be lightweight to minimize players’ dependence on high-performance devices and high-speed networks. Similarly, it is important to avoid substantial hardware demands, which were present in 2 game prototypes that were included in this study: the Instructional remote multi-player VR game requires access to a VR headset, and participants need to own an Xbox 360 device for the Physical Fitness Training Program.

**Consider Time Requirements**

Most of the analog games reviewed in this study (9/15, 60%) were designed for longer durations of gameplay (up to 3 hours); we could only retrieve the predicted duration for 7% (3/43) of the digital games, but the time required to play these games was considerably shorter (up to 15 minutes). The main advantage of short game durations (a few minutes) is that they do not require high levels of commitment, so the games can be played without previous planning. However, they may fail to elicit a flow state, reflection, or both. In contrast, long game durations (several hours) may deter potential users, especially if the game is multiplayer and requires scheduling a time that is appropriate
for all players. Moreover, players may lose interest if a game lasts too long. Therefore, it is important to carefully design pandemic-themed games for optimal duration, depending heavily on the target group, game type, and intended transfer effects.

**Challenge and Progression**

To avoid cognitive overload, it is generally advised that games provide a clear distinction between the active and reflective phases of gameplay. Marne et al [105] suggested that intensive action phases that engage players emotionally and shift their focus toward a game’s goal can be used for practice and training, whereas less intensive phases should be provided for reflection and relief purposes. During reflective phases, games should provide feedback to players so that they can understand the consequences of their actions and keep track of their own progress. However, this information must be provided in a way that does not interfere with the player’s state of flow as this could cause them to lose interest in the game.

A possible solution proposed by Marne et al [105] is that individual items of obtainable knowledge or skills are represented as collectible virtual objects (rewards) that can be showcased or used as an asset later in the game. In addition to material rewards or achieving (publicly displayed) high scores, sensory stimulation (eg, sounds or graphics indicating an achievement) can also be rewarding, as can be messages of affirmation or commendation. For example, games such as *Escape COVID-19* included a positive message at the end of each level to strengthen players’ motivation to comply with guidelines [92]. In such cases, simple, commonly used terminology can help a game more effectively reach broader target audiences.

**Terminology for and Testing of Serious Games**

In addition to the aforementioned considerations, our investigation revealed a need to agree upon a shared terminology and standardized testing approaches to improve the discourse on serious game development. From a scientific point of view, the wide diversity of findings at present renders it practically impossible to make any definitive conclusions or generalizations. A commonly agreed-upon method for effectiveness testing could facilitate the comparison of different approaches in game design and encourage the development of more impactful games. Consequently, we propose that a shared database of knowledge be created, such as the recently developed co.LAB methodological framework, which was implemented in a collaborative web platform that allows for the co-designing, codevelopment, and coevaluation of serious games [118].

**Limitations**

Despite all the efforts to provide a comprehensive review of COVID-19–themed games, we acknowledge that other games that were not included in this report exist (eg, competition and game jam entries were excluded). Furthermore, information on games that were not described in scientific papers was obtained from various web-based sources, potentially limiting the study’s reliability. Nevertheless, we believe that this study correctly reflects the diversity of COVID-19–themed games.

**Conclusions**

This study aimed to provide a structured overview of serious games developed in the context of the COVID-19 pandemic to improve people’s well-being through entertainment and education. We identified 66 diverse games, most of which were digital (43/66, 65%), were developed by research institutions in 2020 (13/66, 20%), and originated in Europe and North America (38/66, 58%). An analysis of the games’ characteristics was performed to identify potential pandemic-specific challenges and opportunities for improvement, and some recommendations were made based on the findings of the reviewed studies and existing game theories.

In total, 2 additional themes emerged as a result of this overview. First, better planning for effective dissemination of the games appears to be necessary if the goal of these games is to reach the broadest audience possible. Second, to increase the impact of similar future interventions, the collective effort of game developers and researchers is needed to advance the discourse on game design and testing.

**Acknowledgments**

This research was supported by Vici grant 453-16-009 of the Netherlands Organisation for Scientific Research, Division for the Social and Behavioural Sciences, awarded to PMAD.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
[DOC File, 111 KB - games_v1i1e41766_app1.doc ]

Multimedia Appendix 2
[DOC File, 162 KB - games_v1i1e41766_app2.doc ]

Multimedia Appendix 3
Intended transfer effects of COVID-19–themed games.

Multimedia Appendix 4

References
19. Howley D. The world is turning to video games amid coronavirus outbreak. Yahoo Finance Home Page. 2020. URL: https://finance.yahoo.com/news/coronavirus-world-turning-to-video-games-150704969.html?guce_referrer=aHR0cHM6Ly9ibG9nLmNvbXNpZmVuY3JhbGljaW50LmNvbS8&guce_referrer_sig=AQAAAAKXHU1OF3HFeK9pEv4LCfWjVvseAu0KCT9oyn_cObt0SSBlvwmOnT3OWV3-4sTvJu8vsoamYVFDFPPHFPLvPTNJ4Panw2SnvGfE3YWy8NpFHpaxLgXaxmV2aoX-UXk2K2BAIsd9qw7.2C/2W82m0rTq5edbC7NxbD-uPldAmm4acbWbb [accessed 2021-12-23]


47. At-home scavenger hunt. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/mentalhealth/stress-coping/parental-resources/
52. You Make Me Sick Board Game. The Partnership In Education, Duquesne University. 2020. URL: https://drive.google.com/file/d/12KZHaJy4Dkf47DcbWhiaL2kefGSB5I/view?usp=sharing [accessed 2021-11-12]
56. Corona Bee game. Focus Games. 2020. URL: https://www.corona-beegame.co.uk/
60. COVID Safety Simulation: CAMPUS LIFE. Arizona State University. 2020. URL: https://xr.asu.edu/exp/covid/ [accessed 2021-11-12]
63. Dino-Store. Digital Integrative Liberal Arts Center, Georgia Institute of Technology, Georgia Tech. 2021. URL: https://lingfengzyc.itch.io/dino [accessed 2021-11-12]
66. GO VIRAL!. 2021. URL: https://www.goviralgame.com/books/go-viral/ [accessed 2021-11-12]
73. Pyson Games Presskit. I Am the Shield. URL: https://iamtheshield.com/presskit/
74. Destroy COVID. URL: https://destroycovidgame.com/
76. Lockdown! an engaging and exciting educational card game. Shoope. URL: https://shoope.sg/Lockdown-an-engaging-and-exciting-educational-card-game-i.205719557.15017303132


A Gamified Real-time Video Observed Therapies (GRVOTS) Mobile App via the Modified Nominal Group Technique: Development and Validation Study

Siti Aishah Abas1*, MBBS, MPH; Nurhuda Ismail1*, MD, MPH, DrPH; Yuslina Zakaria2*, PhD; Ismassabah Ismail3*, BIT, MCS, PhD; Nurul Hidayah Mat Zain4*, BCS, MSCS, PhD; Siti Munira Yasin5*, MbBChBao, MPH, DrPH; Khalid Ibrahim6*, MBBS, MHHSM, CMIA; Asmah Razali7*, MD, MPH; Mas Ahmad Sherkawee Mohd Yusoff8*, MBBS, MPH, DrPH; Norliza Ahmad9*, BSc, MBBS, MPH; Thilaka Chinnayah5*, MBBS, DRM, MPH

1Department of Public Health Medicine, Faculty of Medicine, Universiti Teknologi MARA Sungai Buloh Campus, Sungai Buloh, Malaysia
2Department of Pharmacology, Faculty of Pharmacy, Universiti Teknologi MARA Puncak Alam Campus, Puncak Alam, Malaysia
3Centre of Foundation Studies, Universiti Teknologi MARA Cawangan Selangor, Kampus Dengkil, Dengkil, Malaysia
4Faculty of Computer and Mathematical Science, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasmin, Jasmin, Malaysia
5Tuberculosis and Leprosy Control Sector, Disease Control Division, Ministry of Health, Putrajaya, Malaysia
6TB/Leprosy Disease Unit, Selangor State Health Department, Shah Alam, Malaysia
7TB/Leprosy Disease Unit, Negeri Sembilan State Health Department, Seremban, Malaysia

*all authors contributed equally

Corresponding Author:
Nurhuda Ismail, MD, MPH, DrPH
Department of Public Health Medicine
Faculty of Medicine
Universiti Teknologi MARA Sungai Buloh Campus
Sungai Buloh, 47000
Malaysia
Phone: 60 361267174
Fax: 60 361267234
Email: yuda@uitm.edu.my

Abstract

Background: The success rate of tuberculosis (TB) treatment in Malaysia remains below the recommended World Health Organization target of 90% despite the implementation of directly observed therapy, short-course, a physical drug monitoring system, since 1994. With increasing numbers of patients with TB in Malaysia defaulting on treatment, exploring another method to improve TB treatment adherence is vital. The use of gamification and real-time elements via video-observed therapies in mobile apps is one such method expected to induce motivation toward TB treatment adherence.

Objective: This study aimed to document the process of designing, developing, and validating the gamification, motivation, and real-time elements in the Gamified Real-time Video Observed Therapies (GRVOTS) mobile app.

Methods: The modified nominal group technique via a panel of 11 experts was used to validate the presence of the gamification and motivation elements inside the app, which were assessed based on the percentage of agreement among the experts.

Results: The GRVOTS mobile app, which can be used by patients, supervisors, and administrators, was successfully developed. For validation purposes, the gamification and motivation features of the app were validated as they achieved a total mean percentage of agreement of 97.95% (SD 2.51%), which was significantly higher than the minimum agreement score of 70% (P<.001). Further, each component of gamification, motivation, and technology was also rated at 70% or more. Among the gamification elements, fun received the lowest scores, possibly because the nature of serious games does not prioritize the fun element and because the perception of fun varies by personality. The least popular element in motivation was relatedness, as stigma and discrimination hinder interaction features, such as leaderboards and chats, in the mobile app.

Conclusions: It has been validated that the GRVOTS mobile app contains gamification and motivation elements, which are intended to encourage medication adherence to TB treatment.
video directly observed therapy; VDOT; mobile health; mHealth; tuberculosis; medication adherence; directly observed therapy; video-observed therapy; mobile app; mobile health app; gamification

Introduction

More than 10 million cases of tuberculosis (TB) are reported each year globally [1]. In Malaysia, from 2010 to 2015, TB cases increased from 68.4 to 79.6 cases per 100,000 people [2]. In addition, according to a model projection, the observed and projected TB incidence in Malaysia will reach 300,000 cases in 2030 [3].

In Malaysia, directly observed therapy, short-course (DOTS) is a method to ensure medication compliance by having a trained health care worker or other designated individual provide the prescribed TB drugs and watch the patient take every dose. DOTS, which was implemented in 1994, has resulted in a treatment success rate from 76% in 2013 to 81% in 2017 [2]. However, this rate has remained below the recommended World Health Organization target of 90%. In addition, there is an increase in the prevalence of TB treatment default in Malaysia, which has ranged from 4% in 2010 to 4.8% in 2015 and to 5.6% in the latest study [4]. The increasing number of TB cases indicates that there are still issues and challenges that need to be addressed at all levels.

Specifically, the problems currently facing DOTS can be categorized according to the 3 main stakeholders. For patients with TB, the compulsory need for daily DOTS monitoring results in stigma by the public and absence from home or work responsibilities [5-7]. Health workers and policy makers are hesitant to fully implement DOTS, as there are inadequate human resources, increased TB management costs, less participation from lower management, and a lack of public awareness [8,9].

To address these challenges, the World Health Organization recommends the use of digital technologies to promote TB medication adherence [10]. Video directly observed therapy (VDOT) was introduced to replace physical DOTS and has proven to significantly reduce the cost of managing TB, improve patients’ access to doctors, and be less disruptive to patients’ work and family life [8,11-13]. According to many studies, VDOT is a more cost-effective method that significantly increases patient treatment adherence compared to conventional DOTS [14-18].

Although data on VDOT are becoming increasingly robust, the system has yet to be rigorously evaluated within low- and middle-income countries, especially regarding its feasibility [19]. This is quite worrying, as the implementation of VDOT requires complete access to hardware and internet connectivity, which some countries cannot afford [10]. However, in Malaysia, the number of smartphone users is growing and expected to reach over 33 million by 2024, with an 87.36% smartphone penetration, which suggests that VDOT can be implemented in Malaysia [20].

Making VDOT available via mobile apps could make drug monitoring more convenient and effective [21]. However, with the large numbers of mobile health apps in existence, the problem is often about the sustainability of their use [22]. As a solution, integrating gamification elements inside mobile apps can positively impact health and well-being, improve health behaviors and patient engagement, decrease health care use, and empower patients to self-manage their disease [23,24]. In addition, the use of real-time elements such as virtual reality and augmented reality has proven to increase learning effectiveness and behavior modification, correct medication identification, correct self-administration of medication, and support patient counseling practices [25-27].

Thus, in our Gamified Real-time Video Observed Therapy (GRVOTS) mobile app, the integration of gamification and real-time elements is expected to increase patient motivation. The purpose of this study was to validate the gamification, motivation, and real-time element in GRVOTS, a mobile app for VDOT, from the perspective of the service provider (expert review).

Methods

The mobile app prototype was developed from February 2021 to May 2022. The developmental process as a whole, including content and prototype development as well as content validation (nominal group technique [NGT]), was performed using the design science research process model [28].

Development

Content development of the prototype involved a few stages of literature review, mapping, and justification of the new framework. From the literature review, we identified 3 frameworks that could be used as features of app: gamification framework, video reality and motivation framework, and technology feature framework. The gamification component, as the foundation of the proposed gamified mobile app GRVOTS, will be based on the validated framework for the gamification of diabetes self-management called The Wheel of Sukr [29]. The framework consists of 8 components: fun, esteem, growth, motivation, sustainability, socializing, self-representation, and self-management. For the video reality and motivation framework, we used the attention, relevance, self-representation, and self-management. For the video reality and motivation framework, we used the attention, relevance, confidence, and satisfaction (ARCS) model of motivational design [30]. In terms of the mapping procedure, the ARCS model of motivational design can be combined with the gamification elements to foster motivation [31]. The dynamic nature of gamification, such as self-management, self-representation, and fun, can be equal to satisfaction in the ARCS model. Further, elements of gamification, such as esteem, reward, growth, and socializing, can be equal to the components of confidence in the same model. Subsequently, these 2 frameworks can be integrated where gamification elements are
added to the categories of confidence and satisfaction that are based on the ARCS model. Table 1 shows the categories and subcategories of the proposed model in matrix form.

Game dynamics can improve user desire and motivation by establishing rules that encourage users to explore and learn about the apps [32]. Figure 1 shows a screenshot of the main function of the GRVOTS mobile app and its relation to our intended gamification, motivation, and real-time elements from the framework.

Table 1. Matrix of the ARCS\textsuperscript{a} model of motivational design and The Wheel of Sukr gamification model.

<table>
<thead>
<tr>
<th>Model technique</th>
<th>Category</th>
<th>Relevance</th>
<th>Confidence</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCS (motivation)</td>
<td>Attention</td>
<td>Link previous experience (motive matching)</td>
<td>Self-growth</td>
<td>Reward</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perceive with future usefulness (motive matching)</td>
<td>Learning requirement</td>
<td>Immediate application</td>
</tr>
<tr>
<td>The Wheel of Sukr</td>
<td></td>
<td>—</td>
<td>Success</td>
<td>—</td>
</tr>
<tr>
<td>(gamification)</td>
<td></td>
<td>—</td>
<td>Opportunities</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>—</td>
<td>Personal responsibilities</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>—</td>
<td>Esteem</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>—</td>
<td>Reward</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>—</td>
<td>Growth</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>—</td>
<td>Socializing</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>—</td>
<td>Sustainability</td>
<td>—</td>
</tr>
</tbody>
</table>

\textsuperscript{a} ARCS: attention, relevance, confidence, and satisfaction.

\textsuperscript{b} Not applicable.

Figure 1. The screenshot from the apps and justification of the related element. GRVOTS: Gamified Real-time Video Observed Therapies; VDOT: video directly observed therapy.
Figure 2. Systematic architecture of GRVOTS. Data from patients were collected through the apps and process with predesigned rules. API: application programming interface; GRVOTS: Gamified Real-time Video Observed Therapies; iOS: iPhone Operating System; TB: tuberculosis; VDOT: video directly observed therapy.

**Content Validation via the NGT (Expert Review)**

The NGT is a structured variation of a small-group discussion to reach consensus. Through the agreement of the description of the elements, the NGT was used in this research as a validation tool to evaluate the presence of the gamification and motivation components intended to be used in the app.

**Sample Size**

The NGT is a small-group technique suited to panel sizes of more than 10 people [33,34]. Therefore, there were 11 experts involved in this NGT session. A panel of experts was involved to validate the gamification elements in the GRVOTS app using physical meetings in 3 different settings.

**Study Population Flow**

This study was conducted iteratively in 3 meetings for a more comprehensive evaluation, as illustrated in Figure 3.

The criteria of the experts involved in the group were different according to each group. For the first group, IT experts were experienced and involved in mobile app development for at least 2 years and well versed in the gamification features of mobile apps. The second and third groups were composed of administrative and health care workers who were directly involved in managing patients with TB in the outpatient environment, respectively.

**Figure 3.** Flow of the modified nominal group technique (NGT) meetings.
Study Settings
The 3 NGT meetings were performed in 3 settings: at the university for the meeting with the IT expert panel, at the district health office for the meeting session with the administrative team, and at the TB clinic for the meeting with clinical experts.

Instrument for NGT
The instrument used in the NGT was a questionnaire, and the items were generated from a literature review based on 3 existing models as per the previous mapping. These models were developed into a panel expert checklist, and the questionnaire had 2 parts: Part A asked about the gamification-motivation–rea-time theory, The Wheel of Sukr [35], and the ARCS model, whereas Part B concerned the technology features by Anderson et al [36].

Implementation of NGT
The implementation of NGT involved experts who were selected according to the scope of the study. The workshop was conducted in a face-to-face meeting by a moderator [34]. The NGT workshop lasted approximately 2 hours. Before the workshop, the experts were given a week to use the app. Some of them experienced the task as a supervisor and some as a patient. They were required to send 3-4 VDOT videos. Table 2 shows the basic steps to carry out the NGT process, and Table 3 shows the 5 steps of data analysis for the NGT.

The data analysis process for the NGT was based on the percentage of agreement where an element is accepted when the percentage of agreement is 70% or more [37]. The 1-sample, 2-tailed t test was used to determine whether the mean (SD) percentage of agreement result on the gamification and motivation elements in this app was significantly higher than the percentage of agreement of 70%, with a level of significance (zt error) less than .05. The software used at this stage was SPSS (version 28.0; IBM Corp) and Microsoft Excel.

Table 2. Steps of the specific guide to implement the nominal group technique session.

<table>
<thead>
<tr>
<th>Step</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction to problem statement and explanation</td>
</tr>
<tr>
<td>2.</td>
<td>Silent generation of ideas in writing</td>
</tr>
<tr>
<td>3.</td>
<td>Round robin phase: sharing ideas</td>
</tr>
<tr>
<td>4.</td>
<td>Discussion of ideas</td>
</tr>
<tr>
<td>5.</td>
<td>Voting and ranking</td>
</tr>
</tbody>
</table>

GRVOTS: Gamified Real-time Video Observed Therapies.

Table 3. The steps of data analysis for the nominal group technique (NGT) steps.

<table>
<thead>
<tr>
<th>Step</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ensuring the number of participants (experts) involved in the study</td>
</tr>
<tr>
<td>2.</td>
<td>The formation and calculation of score value is based on the template data analysis of the NGT</td>
</tr>
<tr>
<td>3.</td>
<td>Convert score values into percentage form to obtain the percentage of agreement:</td>
</tr>
</tbody>
</table>

\[
A = \text{the total number of experts and } B = \text{Likert scale used, ie, 5 points} 
\]

| 4. | Determine the acceptance of components and elements based on the percentage of agreement |
| 5. | Determine the positions of the elements according to the percentage of agreement |

Ethics Approval
Ethical approval was obtained from the Universiti Teknologi MARA Ethical Board and Medical Research and Ethics Committee, Malaysia: NMRR-21-1016-58994 (IIR).

No informed consent was taken in this study as the data were only retrieved retrospectively from the database and no identifiers were collected for this study.

Results
Mobile app development with the integration of gamification and real-time elements was performed via a literature review, and the content validation of the component was performed via the modified NGT.

Prototype
The final beta version of the mobile app used Android as the platform. The prototype was designed to be used by 3 users—patients, supervisors, and TB managers—and each user had a different role in the program. This mobile app’s main function is to provide patients with a way to record daily DOTS intake via video (VDOT) as well as medication side effect reporting. After patients log in, they are directed to set up their personal profiles and learn how to use the VDOT medication reminder, which is the main activity. After each VDOT session
is uploaded, it is followed by pop-up motivational quotes as well as the movement of the progress meter indicator.

The accumulated points collected from the progress meter will be translated into badges in the progress report theme. Subsequently, a daily pop-up message will also be the main reminder for the next medication, and patient will be asked if they noticed any adverse reactions to previous medication, with a selection of options concerning their symptoms. The ability to report adverse effects gives patients access to their own medication diaries, which can be reviewed during medical visits. Throughout the TB treatment journey, the progress report theme will help patients track their journey and redeem the internal and external rewards offered.

Every VDOT report and any side effects noted by patients will be verified by the supervisor as a feedback interaction. This feature enables supervisors to regularly check and understand the patients’ progress as well. In addition, patients can always go to other main theme of “knowledge” to continuously learn more about TB treatment and nutrition.

The app provides different users with access to different functions and main menus. In the patients’ main menu, there are 4 main themes for VDOT. Information regarding the app, progress report, and side effect reporting is shown in Figure 4. For the supervisors’ main menu, there are 2 menus, one to validate the VDOT report and another to validate the side effects, as shown in Figure 5.

Figure 4. Screenshots of the GRVOTS app for patients. GRVOTS: Gamified Real-time Video-Observed Therapies.
NGT Output for Content Validation

A panel of 11 experts were involved in the modified NGT, including 2 gamification experts, 2 public health experts, 1 chest physician, 2 medical officers, 3 medical assistants, and 1 staff nurse. They were asked to rank the app’s elements based on the expert panel checklist. Tables 4-6 show the results and rankings of gamification, motivation, and technology feature elements via the NGT.

The total percentage of agreement among experts was 97.95% (SD 2.51%), which was significantly higher than the minimum of 70%, with a difference of 27.947% (95% CI 26.74%-29.15%; P<.001). The t test confirmed the validity of all gamification and motivation components in the app.
Table 4. Results of the nominal group technique for gamification elements.

<table>
<thead>
<tr>
<th>No</th>
<th>Main element</th>
<th>Explanation element</th>
<th>Percentage of agreement, mean (SD)</th>
<th>Acceptance result</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Self-management</td>
<td>• This app has a dashboard with a progress meter, medication reminders, and a VDOT(^\text{a}) feature to help patients self-manage their medication.</td>
<td>98 (2.22)</td>
<td>Accepted</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This app has adverse effect reporting to help patients self-manage their reactions to medications.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Self-representative (avatar)</td>
<td>• This app has a personalized name on the dashboard along with the roles of the user.</td>
<td>97 (0.00)</td>
<td>Accepted</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Fun</td>
<td>• This app has a progress meter, a progress bar, motivational quotes, and badge rewards that add a fun element for the patient.</td>
<td>96 (0.8)</td>
<td>Accepted</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Esteem</td>
<td>• This app has a progress meter, a progress bar, motivational quotes, and a feedback mechanism aimed at boosting patients’ self-esteem.</td>
<td>99 (1.386)</td>
<td>Accepted</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Growth</td>
<td>• This app has a feedback mechanism that helps patients with self-growth.</td>
<td>100 (0.00)</td>
<td>Accepted</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This app empowers patients’ self-growth with medication information.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sustainability (trigger)</td>
<td>• This app has medication reminders aimed at the sustainability of its use.</td>
<td>97 (0.00)</td>
<td>Accepted</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Motivation</td>
<td>• This app has a progress meter, a progress bar, and badge rewards that boosts patient motivation.</td>
<td>98 (2.358)</td>
<td>Accepted</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>Socializing</td>
<td>• This app has a feedback mechanism that enables patient-supervisor communication.</td>
<td>100 (0.00)</td>
<td>Accepted</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^{a}\)VDOT: video directly observed therapy.

Table 5. Results of the nominal group technique for motivation elements.

<table>
<thead>
<tr>
<th>No</th>
<th>Main element</th>
<th>Explanation element</th>
<th>Percentage of agreement, mean (SD)</th>
<th>Acceptance result</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Attention</td>
<td>• This app has medication reminders, a VDOT(^{a}) feature, and adverse effect reporting to hold patients’ attention and achieve active participation.</td>
<td>99 (1.869)</td>
<td>Accepted</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This app has a progress meter that alerts patients regarding the status of their performance.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This app has a performance indicator (progress bar) that alerts patients about medication compliance over time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Relevance</td>
<td>• This app has a VDOT feature that links patients’ current remote and previous in-person experience of taking medication.</td>
<td>98 (0.00)</td>
<td>Accepted</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Confidence</td>
<td>• This app has motivational quotes to boost patient confidence.</td>
<td>100 (0.00)</td>
<td>Accepted</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Satisfaction (achievement)</td>
<td>• This app has a VDOT feature, adverse effect reporting, a progress meter, a progress bar, motivational quotes, and badge rewards that help patients feel satisfied with their daily medication compliance.</td>
<td>99 (1.361)</td>
<td>Accepted</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^{a}\)VDOT: video directly observed therapy.
Table 6. Results of the nominal group technique for technology feature elements.

<table>
<thead>
<tr>
<th>No</th>
<th>Main elements</th>
<th>Explanation element</th>
<th>Percentage of agreement, mean (SD)</th>
<th>Acceptance result</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Automation</td>
<td>• This app has automated medication reminders, pop-up notifications, and reminders to both the patient and supervisor that made it easy to use.</td>
<td>97 (0.635)</td>
<td>Accepted</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Ease of use (automation)</td>
<td>• This app has automated pop-up notifications to remind the supervisor to approve the adverse effect reports.</td>
<td>100 (0.00)</td>
<td>Accepted</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Real-time feature</td>
<td>• This app has a real-time stamped VDOT(^a) feature.</td>
<td>89 (0.00)</td>
<td>Accepted</td>
<td>7</td>
</tr>
</tbody>
</table>
| 4  | Self-guided video | • This app has a self-guided VDOT tutorial video that can benefit the patient.  
   • This app has self-guided adverse effect symptom choices that can help patients identify symptoms.  
   • This app has a self-guided feature to provide the patient with knowledge regarding nutrition and medication.  
   • This app has a good tutorial on how to use the app. | 99 (1.361)                         | Accepted         | 3       |
| 5  | Aesthetic     | • This app has good aesthetic features.                                               | 97 (0.00)                         | Accepted         | 6       |
| 6  | Navigation    | • This app has a good flow of navigation from one page to another.                   | 100 (0.00)                        | Accepted         | 2       |
| 7  | Tactile feedback | • This app has good tactile or haptic feedback.                                      | 98 (0.00)                         | Accepted         | 4       |

\(^a\)VDOT: video directly observed therapy.

Discussion

This comprehensive GRVOTS mobile app was developed based on the integration of gamification and real-time motivational elements of autonomy, competence, relatedness, attention, relevance, self-control, confidence, and satisfaction. These components were successfully validated and significant.

Mobile App Developed

Our GRVOTS mobile app was able to motivate patients toward medication adherence using the ARCS model of motivation, which was translated via its gamification and motivation features. First, the interactive VDOT feature can stimulate the patient’s feelings of attention and arousal. Second, by linking the previous experience of medication intake physically with the experience of using GRVOTS, the element of “relevance” can also be instilled in the patient. When patients perceive a high sense of relatedness, they are more likely to exhibit higher engagement with a program [38]. By providing goals along the journey toward medication adherence, the app can instill confidence and self-belief, especially when the patients can digitally see their previous VDOT report and a progress meter of treatment success. As they use GRVOTS to monitor their medication intake, patients experience more freedom of choice and self-control, and the results of treatment adherence will subsequently reinforce the app’s value. The last element of “satisfaction” can also be realized when patients comply with the use of GRVOTS and receive rewards in the form of badges and progress meters.

Content Validation via the Modified NGT

The results of the modified NGT showed that all the components obtained from the literature and related model were validated by the experts during the NGT session, as shown in Tables 4-6. Based on the results, the elements were prioritized based on the percentage of acceptance. Considering these rankings, some of the GRVOTS functions were improved to provide a better mobile app for the next pilot study.

According to the results for the gamification elements, the growth element received the highest scores, whereas fun received the lowest scores. By definition, “fun” aspects within the gamification elements in this GRVOTS app are exemplified by a progress bar, a progress meter, inspiration quotes, and badge rewards. The fun element was voted as a less visible element in this app, perhaps due to many factors, such as the perception of fun, which varies for different people, as well as the nature of serious games, which does not prioritize the fun element. A number of studies have described that the perception of gamification features differs according to the gamer type, gender, and personality of the player. For instance, extroverts like rewards and leaderboards, which appear to be more entertaining, but introverts prefer badges and feedback [39-41]. Health apps are also considered serious games that are played for purposes other than pure entertainment [42]. In this case, having a fun component is not a priority. According to a study regarding serious games, other elements, such as explicit learning tasks, instruction, and support built into the game or added by teachers, may be more important than having fun while playing [43].
For the motivation elements of the ARCS model, the highest ranked element was confidence and the lowest was relevance. It is suggested that when patients perceive a high sense of relatedness, they are more likely to exhibit higher engagement with the program [38]. In mobile apps, relatedness is seen as a feature of leaderboards or chats that encourages engagement and collaboration to achieve a particular objective [41]. However, in this GRVOTS mobile app, the relevance element is evidenced by allowing daily communication between patients and supervisors via VDOT and adverse event reporting components without the leaderboard or chat features. This is because most of our patients with TB refused to disclose their condition and communicate with other patients due to the stigma associated with the disease.

For the technology feature elements, the highest ranked was ease of use and the lowest was real-time features. There are 2 type of VDOT: live VDOT, also known as synchronous VDOT, in which patients and providers interact in real time [12,44]; as well as asynchronous technologies that record, upload, and digitally store videos for future review [11,19,45]. Synchronous VDOT has the advantage of human interaction but is not a feasible option, as patients and workers need to find time to meet. Asynchronous VDOT is more flexible but can be manipulated easily by sending the same recorded video. Thus, apps with real-time features ensure the originality of the video and simultaneously generate a greater degree of user engagement [25-27]. Although the real-time feature in our app provides details concerning the time and date of the video, feedback from the experts indicated that we should time stamp the videos so that they can be identified more easily later. Since users found it challenging to determine if their VDOT session was successfully uploaded or not, the addition of an upload bar was also recommended.

Strengths and Limitations
The GRVOTS mobile app can benefit users in many ways. For patients with TB, the app can help patients gain self-control, boost their self-esteem, and motivate them to take their medications. For health care workers, this app made it easier for TB system management to detect DOTS defaulters and manage them accordingly. As a portable device, mobile apps enable monitoring that can be done anywhere, saving money and time and boosting patient engagement with the DOTS program. In summary, this app can also initiate a patient self-care system and reduce dependency on health care providers such as doctors and nurses. The limitation of this study is that the mobile app prototype is only being developed for the Android platform because of time and logistics. The app does not have a virtual reality feature, and improvements are needed in the future. This study conducted the validation of the gamification and motivation elements only from the perspective of the experts and not from the perspective of patients, which may limit the review’s validity; an analysis regarding its usability among patients will be conducted in the future.

Recommendation
In the future, GRVOTS should also be available on other platforms, especially the iPhone Operating System. The language options should also include English, Chinese, and Tamil, as these languages are frequently used in multiethnic communities in Malaysia. The use of various languages will expand the benefits to more users. This will enhance knowledge transfer and improve users’ understanding. Next, a usability study to access the user experience will be conducted, followed by an effectiveness study via a single-arm intervention study, in which patients will use the app for DOTS during intensive phase up to 2 months monitoring, followed by an assessment at 3 time intervals to evaluate their medication adherence, motivation, and the usability of the app.

Conclusion
More comprehensive and efficient TB system management via VDOT mobile app monitoring is a way to improve patient treatment adherence. According to the literature review, gamification elements can motivate patients; thus, by integrating the uniqueness of gamification and motivation elements in an app, gamification will increase patient motivation, ensure the sustainability of use, and ultimately increase patient adherence. In addition, our GRVOTS mobile app connects up to 3 users (eg, patients, supervisors, and administrators) remotely and enable DOTS monitoring to be performed from anywhere. Based on the study findings, the GRVOTS mobile app has been validated by the expert panel as having the intended elements of gamification, real time, and motivation. Next, a usability study of the GRVOTS mobile app will be conducted to measure the user experience among patients, followed by a single-arm intervention study to assess the app’s effectiveness in increasing patient motivation and medication adherence in TB treatment.

Acknowledgments
The author would like to thank to the Director General of Health Malaysia for his permission to publish this study. We give many thanks also to the Universiti Teknologi MARA.

This study is funded by the Joint TDR/World Health Organization Western Pacific Region Small Grant Scheme for Implementation Research.

Data Availability
All data supporting the study findings are within the manuscript. Additional details information and raw data are available from the corresponding author on reasonable request.

https://games.jmir.org/2023/1/e43047
Authors' Contributions
SAA and NI conceived the project. SAA contributed to data collection. SAA, NI, and YZ contributed to data analysis. SAA, II, and NHMZ contributed to mapping the app architecture. NI supervised the study. The manuscript was written mainly by SAA with the approval of all authors. All authors read and approved the final version of manuscript for publication.

Conflicts of Interest
None declared.

References


Abbreviations

ARCS: attention, relevance, confidence, and satisfaction
DOTS: directly observed therapy, short-course
GRVOTS: Gamified Real-time Video Observed Therapies
NGT: nominal group technique
TB: tuberculosis
VDOT: video directly observed therapy

©Siti Aishah Abas, Nurhuda Ismail, Yuslina Zakaria, Ismassabah Ismail, Nurul Hidayah Mat Zain, Siti Munira Yasin, Khalid Ibrahim, Asmah Razali, Mas Ahmad Sherzawee Mohd Yusoff, Norliza Ahmad, Thilaka Chinnayah. Originally published in JMIR Serious Games (https://games.jmir.org), 07.03.2023. 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.
Evaluating the Utility of a Psychoeducational Serious Game (SPARX) in Protecting Inuit Youth From Depression: Pilot Randomized Controlled Trial

Yvonne Bohr1,2, PhD; Leah Litwin1,2, PhD; Jeffrey Ryan Hankey1,2, PhD; Hugh McCague3, PhD; Chelsea Singoorie4; Mathijs F G Lucassen5,6, PhD; Matthew Shepherd6, DClinPsy; Jenna Barnhardt1, BSc

1LaMarsh Centre for Child and Youth Research, York University, Toronto, ON, Canada
2Department of Psychology, Faculty of Health, York University, Toronto, ON, Canada
3Institute for Social Research, York University, Toronto, ON, Canada
4Nunabox, Iqaluit, NU, Canada
5Faculty of Wellbeing, Education and Language Studies, The Open University, Milton Keynes, United Kingdom
6School of Psychology, Massey University, Auckland, New Zealand

Corresponding Author:
Jeffrey Ryan Hankey, PhD
LaMarsh Centre for Child and Youth Research
York University
4700 Keele Street
Toronto, ON, M3J1P3
Canada
Phone: 1 6475616464
Email: jrhankye@yorku.ca

Abstract

Background: Inuit youth in Northern Canada show considerable resilience in the face of extreme adversities. However, they also experience significant mental health needs and some of the highest adolescent suicide rates in the world. Disproportionate rates of truancy, depression, and suicide among Inuit adolescents have captured the attention of all levels of government and the country. Inuit communities have expressed an urgent imperative to create, or adapt, and then evaluate prevention and intervention tools for mental health. These tools should build upon existing strengths, be culturally appropriate for Inuit communities, and be accessible and sustainable in Northern contexts, where mental health resources are often scarce.

Objective: This pilot study assesses the utility, for Inuit youth in Canada, of a psychoeducational e-intervention designed to teach cognitive behavioral therapy strategies and techniques. This serious game, SPARX, had previously demonstrated effectiveness in addressing depression with Māori youth in New Zealand.

Methods: The Nunavut Territorial Department of Health sponsored this study, and a team of Nunavut-based community mental health staff facilitated youth’s participation in an entirely remotely administered pilot trial using a modified randomized control approach with 24 youths aged 13-18 across 11 communities in Nunavut. These youth had been identified by the community facilitators as exhibiting low mood, negative affect, depressive presentations, or significant levels of stress. Entire communities, instead of individual youth, were randomly assigned to an intervention group or a waitlist control group.

Results: Mixed models (multilevel regression) revealed that participating youth felt less hopeless ($P=0.02$) and engaged in less self-blame ($P=0.03$), rumination ($P=0.04$), and catastrophizing ($P=0.03$) following the SPARX intervention. However, participants did not show a decrease in depressive symptoms or an increase in formal resilience indicators.

Conclusions: Preliminary results suggest that SPARX may be a good first step for supporting Inuit youth with skill development to regulate their emotions, challenge maladaptive thoughts, and provide behavioral management techniques such as deep breathing. However, it will be imperative to work with youth and communities to design, develop, and test an Inuit version of the SPARX program, tailored to fit the interests of Inuit youth and Elders in Canada and to increase engagement and effectiveness of the program.

Trial Registration: ClinicalTrials.gov NCT05702086; https://www.clinicaltrials.gov/ct2/show/NCT05702086
KEYWORDS

psychoeducation; cognitive behavioral therapy; Inuit youth; Nunavut; depression; suicide; resilience; serious game; youth; mental health; teen; adolescent; pilot study; community

Introduction

Inuit are Indigenous Peoples of the Arctic and subarctic regions of Greenland, Alaska, and Canada. In Canada, they are recognized as a distinctive group of Indigenous Canadians—alongside First Nations and Métis—primarily occupying Labrador, Northern Quebec, the Northwest Territories, and Nunavut. Pervasive youth suicide is the most urgent challenge facing Inuit in Canada today, a public health emergency of epidemic proportions [1,2]. Death by suicide in Inuit communities has spiked over the last century [3,4], currently at 9 times the national average and with male youth aged 15-24 being the most affected [5]. Under the fallout of colonialism, Inuit youth in Canada face more mental health challenges than non-Indigenous populations, due to cross-generational impacts of residential schools and multiple continuing systemic inequities [6,7]. The government era of the 1950-60s and beyond, during which the Canadian government forced Inuit into aggregated settlements, imposed a welfare state, and separated Inuit children from their families, has had lasting effects on relationships within family and community, strained kinship across generations and between genders, and left generational gaps in mental health–supporting communications between youth and their parents and Elders [3]. Further, Indigenous youth in general tend to be among those most reluctant to seek help when experiencing distress [8]. This lack of help-seeking may be due in part to a struggle to recognize symptoms of depression, which may present as anger in Inuit youth [4,9].

Those living in remote Northern communities often go without access to the mental health supports available elsewhere in Canada [10] for a number of reasons. First, few health care professionals train or practice in the Canadian North, and frontline workers often experience “burnout” due to excessive workload and personnel shortages [11]. As a result, youth in Nunavut have limited access to specialized services and are often sent out of their communities when they experience severe mental health symptoms [10,12-14]. Finally, adolescents in small, remote communities are especially prone to avoiding professional health services, as they may fear compromised confidentiality due to a lack of privacy and the stigma associated with mental illness. One study found that of all youth who committed suicide in Nunavut in the past decade, only 12% had received medication and 17% had been hospitalized in relation to their mental health, while 89% had not received any mental health support at all [14,15]. Given the reality that there are relatively few frontline mental health workers available in Nunavut communities and stigma impeding access to existing services, it is imperative to identify and develop preventive interventions for depression that are not only effective and culturally appropriate, but also easily accessible [11].

To address the dearth of available in-person mental health support options, this study aims to examine the usefulness of an e-intervention, SPARX [16,17], in Nunavut. SPARX (Smart, Positive, Active, Realistic, X-Factor thoughts) is a psychoeducational serious game (an e-intervention that utilizes gaming for serious purposes) [18] that teaches established cognitive behavioral therapy (CBT) strategies and techniques across 7 levels or modules (see Figure 1 for an example). The game is designed to address depressive symptoms in youth by helping them cope with negative thoughts and feelings, represented in the game as GNATs—gloomy negative automatic thoughts (Figure 2) [16]. SPARX was originally designed and developed at the University of Auckland with the specific needs of certain underserved groups of youth in mind, including Māori Rangatahi, the Indigenous young people of Aotearoa, New Zealand. SPARX’s development included a Māori co-creator (MS), input from Māori CBT experts, and cultural guidance from a kaumatua (ie, respected Elder). Moreover, the game development company was led by a Māori woman and the voice actor (virtual therapist/instructor) was a Māori; additionally, the Māori symbolism, presented in a fantasy format, was woven throughout the intervention [17]. With these very specific cultural foundations in mind, the goal of this pilot study was to evaluate the effectiveness of the original version of SPARX in boosting resilience against depression among Inuit youth in Canada in 11 Nunavut communities. We aimed to assess whether Inuit youth who completed the SPARX program experienced increased resilience and showed a decrease in risk factors related to depression, specifically cognitive distortion, emotional dysregulation, hopelessness, rumination, self-blame, other-blame, and catastrophizing [19].

The Nunavut Territorial Department of Health sponsored this study, and a team of Nunavut-based community mental health staff facilitated youth’s participation in an entirely remotely administered pilot trial. A total of 48 youths from 11 Nunavut communities completed an evaluation of the utility of SPARX in modifying dysfunctional cognitions, reducing symptoms of depression, and enhancing resilience. All youth completed conventional pre- and postintervention measures assessing their current mental health status. Measures were selected based on their use in earlier effectiveness studies [16,17]. Quantitative and qualitative data were collected and used to evaluate the program’s preliminary effectiveness based on the following hypotheses:

- Youth who completed the SPARX program were expected to experience a decrease in depressive symptoms as well as risk factors related to depression, as measured by the Centre for Epidemiologic Depression Scale-Revised (CESD-R), the Hopelessness Scale for Children (HSC), and the Cognitive Emotion Regulation Questionnaire-Short (CERQ-Short; see the “Measures” section).
Youth who completed the SPARX program were expected to experience an increase in factors related to resilience, as measured by the shortened, 12-item version of the Child and Youth Resilience Measure (CYRM-12).

**Figure 1.** Screenshot of a SPARX CBT module. Used with permission of the copyright owner © Auckland UniServices Limited.

**Figure 2.** Screenshot of approaching GNATs. Used with permission of the copyright owner © Auckland UniServices Limited.

### Methods

#### Participants

All 25 Nunavut communities were initially recruited to participate in this study; 11 communities (Baker Lake, Cambridge Bay, Grise Fiord, Hall Beach, Igloolik, Kugaaruk, Kugluktuk, Naujaat [Repulse Bay], Taloyoak, Qikiqtarjuaq, and Resolute) completed the pilot trial. The remaining 14 communities did not complete data collection because of an inability to recruit youth, because there was no local adult facilitator able to support youth, or because facilitators had to leave the project due to extraneous circumstances. Communities were chosen based on feasibility and required both an interested community facilitator with the time to commit to SPARX and youth who were interested in the project. As many as 3-4 youths were recruited in each community with a projected participant count of 40.

Youth participants were between the ages of 13 and 18 and had been identified by the community facilitators as exhibiting low mood, negative affect, depressive presentations, or significant levels of stress. Youth had to demonstrate sufficient English language comprehension to use and understand SPARX.
Although such an approach is not reflective of Indigenized research generally, these inclusion criteria were suggested by local facilitators due to the heavy use of the English language within the SPARX game. Initially, youth were excluded if they showed limited cognitive abilities, psychotic presentation, severe depression, or elevated suicide risk, or if they were currently receiving or had previously received CBT, interpersonal therapy, or antidepressant medication within the past 3 months. However, due to safety and ethics concerns about vulnerable youth feeling left out, at the discretion of local mental health experts on the team, these exclusion criteria were relaxed. The 25 facilitators were appointed by staff at the Department of Mental Health, Government of Nunavut. In total, 24 of 48 youths who initially signed up completed the study (see the “Design” section for attrition rates at each time point).

Procedure
A preproject planning meeting was held in Toronto with managers of the Territorial Department of Health, a team of community mental health staff from Nunavut, and potential on-site facilitators on March 27, 2014. Procedural aspects of the SPARX pilot project were based on contributions and feedback from attendees of this meeting. The meeting focused on the importance of the SPARX project in Nunavut communities, how the SPARX project could achieve cultural appropriateness, and what traditional and cultural knowledges must be incorporated to gain youth trust and participation. In a collaboration among Nunavut Department of Mental Health staff, community stakeholders, and the coordinating team at York University, the design described in this work was adopted at the planning meeting.

Ethics Approval
Approval for the SPARX pilot was obtained through the Human Participants Sub-Committee of the York University Office of Research Ethics (York University HPRC Certificate No. 2015-070). Approval was also obtained from the university’s Advisory Group for Research Involving Indigenous Peoples. This subcommittee is guided by Chapter 9 of the Tri-Council Policy Statement “Research involving the First Nations, Inuit, and Métis Peoples of Canada,” and its own Guidelines for Research Involving Indigenous Peoples. In addition, the research team received approval in the form of a research license from the Nunavut Research Institute, which represents all 25 Nunavut communities (license number 02 004 15N-M).

Design
A modified randomized control approach was adopted for this trial. Community facilitators, responsible for recruiting youth and facilitating gameplay, were trained prior to youth recruitment and played SPARX to familiarize themselves with the game. Once training was complete, community facilitators recruited youth, secured consent, and administered preintervention measures. Entire communities, instead of individual youth, were randomly assigned to an intervention group (group A) or a waitlist control group (group B), which is why a modification was made to the randomized control trial, as described in the following section.

The sequence of SPARX play differed for the youth depending on whether they were in group A or group B (Figure 3). Both groups completed the preintervention surveys at time 1, the beginning of the study. Group A youth then began to play SPARX for 7 weeks while group B waited. During their wait time of 7 weeks, group B youth were not required to participate in any SPARX activities or meet with the community facilitator, and they were provided with no additional SPARX-related information. After 7 weeks (time 2), group A youth completed postintervention surveys, while group B youth began their engagement with SPARX by first completing an additional set of preintervention surveys, immediately followed by 7 weeks of SPARX play. Once group B youth completed their SPARX gameplay (time 3), they next completed postintervention surveys. Attrition rates were 35% (7/20) for group A (the intervention group) at time 2, 32% (9/28) for group B (the waitlist control group) at time 2, and a further 47% (9/19) for group B at time 3. This was due to high community facilitator turnover, timing of holidays, loss of interest, and arising community crises that took facilitators away from supporting SPARX usage. Similar attrition rates are not uncommon in pilot studies in remote communities.

Youth played SPARX on the laptops in either the community facilitator’s office or in an office in their school. Youth were asked to engage in SPARX play once a week; however, due to complications in some communities (eg, crises, snowstorms), some youth played more or less frequently than this. Youth engaging in SPARX were always supported and accompanied by a community facilitator, given the exploratory nature of this novel study. This assistance was provided to ensure that any support youth needed with gameplay or with their mental health difficulties was prudently available while this new intervention was being piloted. The sequence of SPARX play is detailed in the following sections.
Measures
Pre- and postintervention surveys consisted of the same set of measures at both time points. Measures were either completed online via SurveyMonkey or in printed paper format. Community facilitators were on-site with the youth while they completed the measures to provide support such as reading questions out loud to the youth, or talking through answers if any youth felt vulnerable, at risk, or triggered by the content of the questions. Measures were selected based on their relevance; use in earlier, related studies; established psychometric properties for youth; sensitivity to change over time; ease of administration; length and clarity; and constraints based on funding sources.

The goal of this pilot study was to replicate, as best as possible in this new context, the studies conducted with Māori youth [16] to assess the potential usefulness of SPARX as an intervention with Inuit youth in Canada. It was clear to the researchers that should funding subsequently be obtained for a more comprehensive study, more culturally appropriate questionnaires would then be developed, together with Inuit communities, to assess youth depression and resilience [20]. As of the time of the writing of this paper, that has indeed happened, and a study of a culturally adapted version of SPARX for Inuit youth is underway [21]. In the absence of culturally appropriate outcome measures to assess mental health factors related to suicidality in Inuit youth, and to remain as consistent as possible with existing studies of SPARX conducted in New Zealand, the following primary outcome measures were chosen to assess changes in depression, hopelessness, resilience, and cognitive emotion regulation within these communities.

The CES-D-R [22] is a succinct, 20-item, self-report scale that aims to measure current depressive symptomatology. Each item is a symptom associated with depression. The CES-D-R has acceptable internal consistency (Cronbach α = .85) and test-retest reliability [23]. Validity was also established based on correlations with other assessment tools [24]. A sample item includes, “In the past week, I was bothered by things that don’t usually bother me.” Although it has not yet been evaluated within Inuit communities, the CES-D-R is the most commonly used measure of depression among Indigenous youth, and has demonstrated a good model fit, with a root-mean-square error of approximation [25] of 0.6.

The HSC [26] contains 17 true-false items, which describe negative future expectations and negative present attitudes. Scores range from 0 to 17. Internal consistency is α = .97 and test-retest reliability is r = .52. The HSC correlated positively with depression (r = .58) and negatively with self-esteem measures (r = -.61) and social skills (r = .39) [26]. A sample item includes, “I might as well give up because I can’t fix it without hurting myself or other people (for example hitting others or saying nasty things).” The CYRM-12 has frequently been used with Indigenous populations, showing good content and internal validity, but has not been evaluated with Inuit communities [28].

Data Analysis
Because of the challenges encountered with youth recruitment, and the knowledge that an ideal sample size would likely be unattainable, no formal power calculations were conducted prior to beginning this study. To analyze mental health changes in Inuit youth in Canada over time, as measured through the 4 questionnaires, paired samples t tests (both 1- and 2-tailed—see below) comparing pre- and postintervention outcomes and repeated measures ANOVAs were conducted via SPSS Statistics version 26 (IBM Corp). ANOVAs allowed for examining the effect of time interacting with the intervention group or the waitlist control group. Paired samples t tests and ANOVAs pooled group A and group B data to account for missing data across groups and fewer completed postmeasures from group B youth. As a result of pooled data, there was some loss in accuracy due to the omission of select cases that did not fit the structure required by these simpler statistical methods.

Given the structure of the study—the comparison of the intervention group (group A) and the waitlist group (group B) at all (up to 3) time points for the pre- and postmeasures (group A: premeasure at time 1, postmeasure at time 2; group B: premeasure at time 1, premeasure at time 2, and postmeasure at time 3)—mixed models (multilevel regression) using pooled data were conducted via SPSS software. Multilevel regression models were appropriate given their ability to detect individual differences across different treatment options (time points for youth in group A vs group B). Results of the multilevel regression models are considered the most accurate as they use all the nonmissing data (important with a small sample size), reflect the full design of the study, and allow for comparison between group A and group B at different time points. Multilevel regression models were fit with (1) a randomly varying intercept for each youth to account for their differing initial levels of personally identified health and (2) an intervention indicator.
variable (0 for not having started playing the SPARX program vs 1 for completing the treatment of playing SPARX) to account for the effect of the treatment (ie, engaging in SPARX). A time variable was not needed as there were at most 2 (group A) to 3 (group B) measurement occasions with a horizontal, time-independent trajectory after accounting for the effect of the intervention. The assumptions of the mixed models were checked (eg, normality of the residuals) and found reasonable.

Researcher Characteristics
Enduring colonial approaches to research in Canada have severely diminished Inuit sovereignty with respect to research activity in the Inuit homeland [30]. In the spirit of accountability and reconciliation, we believe it is important to acknowledge that the research team is diverse and consists of Indigenous and non-Indigenous team members. CS, our Inuit community liaison, is an Inuk and a longstanding research partner with YB and her team of researchers at York University. MS is a Māori researcher who co-developed the SPARX program with Indigenous youth in New Zealand. JB is an Indigenous status First Nations student, born and raised in Tyendinaga/Kenhteke Mohawk Territory. YB, LL, JRH, HM, and MFGL are White researchers from the South (Qallunaat).

Results
Overview
Table 1 provides a summary of the baseline and postintervention results for participants in groups A and B. Mean scores and the SDs for the 4 outcome measures (by subscale for the CERQ-Short) are provided below.

Table 1. Descriptive statistics by group and time point.

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Group A time 1 (n=20), mean (SD)</th>
<th>Group A time 2 (n=13), mean (SD)</th>
<th>Group B time 1 (n=28), mean (SD)</th>
<th>Group B time 2 (n=19), mean (SD)</th>
<th>Group B time 3 (n=10), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYRM-12:&lt;sup&gt;a&lt;/sup&gt;: Resilience</td>
<td>44.70 (8.04)</td>
<td>41.07 (10.22)</td>
<td>43.32 (7.54)</td>
<td>43.62 (9.44)</td>
<td>45.30 (8.42)</td>
</tr>
<tr>
<td>HSC:&lt;sup&gt;b&lt;/sup&gt;: Hopelessness</td>
<td>4.52 (2.75)</td>
<td>3.79 (3.09)</td>
<td>5.39 (2.50)</td>
<td>4.73 (3.34)</td>
<td>3.67 (2.58)</td>
</tr>
<tr>
<td>CESD-R:&lt;sup&gt;c&lt;/sup&gt;: Depression</td>
<td>22.90 (8.30)</td>
<td>23.21 (8.86)</td>
<td>25.50 (8.54)</td>
<td>23.68 (8.58)</td>
<td>21.20 (7.60)</td>
</tr>
<tr>
<td>CERQ-Short subscale 1: Self-blame</td>
<td>4.90 (2.70)</td>
<td>5.00 (1.84)</td>
<td>6.04 (2.13)</td>
<td>6.10 (2.13)</td>
<td>4.20 (1.75)</td>
</tr>
<tr>
<td>CERQ-Short subscale 2: Acceptance</td>
<td>6.30 (2.24)</td>
<td>5.64 (2.20)</td>
<td>7.39 (2.33)</td>
<td>7.26 (2.37)</td>
<td>7.90 (1.73)</td>
</tr>
<tr>
<td>CERQ-Short subscale 3: Focus on Thought/Ruminating</td>
<td>5.70 (2.39)</td>
<td>5.07 (1.90)</td>
<td>7.07 (2.16)</td>
<td>6.68 (2.29)</td>
<td>6.00 (1.89)</td>
</tr>
<tr>
<td>CERQ-Short subscale 4: Positive Refocusing</td>
<td>5.60 (1.79)</td>
<td>5.21 (2.19)</td>
<td>5.29 (1.90)</td>
<td>5.58 (2.27)</td>
<td>6.40 (2.07)</td>
</tr>
<tr>
<td>CERQ-Short subscale 5: Refocus on Planning</td>
<td>7.10 (2.40)</td>
<td>5.86 (2.51)</td>
<td>6.50 (2.23)</td>
<td>6.58 (2.24)</td>
<td>6.10 (2.42)</td>
</tr>
<tr>
<td>CERQ-Short subscale 6: Positive Reappraisal</td>
<td>7.40 (2.35)</td>
<td>7.07 (2.30)</td>
<td>8.00 (1.61)</td>
<td>7.68 (1.82)</td>
<td>6.70 (2.47)</td>
</tr>
<tr>
<td>CERQ-Short subscale 7: Putting Into Perspective</td>
<td>6.40 (2.39)</td>
<td>5.79 (2.01)</td>
<td>6.43 (2.28)</td>
<td>6.32 (1.60)</td>
<td>5.50 (2.37)</td>
</tr>
<tr>
<td>CERQ-Short subscale 8: Catastrophizing</td>
<td>5.50 (2.70)</td>
<td>4.29 (1.07)</td>
<td>6.50 (2.46)</td>
<td>5.89 (2.51)</td>
<td>5.60 (2.41)</td>
</tr>
</tbody>
</table>

Hypothesis One: Youth Who Completed the SPARX Program Were Expected to Experience a Decrease in Depressive Symptoms and Risk Factors Related to Depression, as Measured With the CESD-R, HSC, and CERQ-Short Questionnaires
Participating youth reported feeling less hopeless, and engaged in less self-blame, rumination, and catastrophizing following the SPARX intervention (Table 2). The intervention was a statistically significant predictor of the HSC score ($b$=–1.08; $t_{51.2}$=–1.95; $P$=.03<.05; 1-tailed). The intervention tended to decrease the CERQ-Short subscale score 1: Self-blame by close to 1 (0.86) unit. The intervention was a statistically significant predictor of the CERQ-Short subscale score 3 (Focus on Thought/Ruminating; $b$=–0.77; $t_{53.6}$=–1.84; $P$=.03<.05; 1-tailed). The intervention was a statistically significant predictor of the CERQ-Short subscale score 8 (Catastrophizing; $b$=–0.90; $t_{52.7}$=–2.00; $P$=.02<.05; 1-tailed). For scores on the CESD-R, and the CERQ-Short subscale scores 2, 4, 5, 6, and 7 (Acceptance, Positive Refocusing, Refocus of Planning, Positive Reappraisal, Putting Into Perspective), the intervention indicator variable at the $\alpha$=.05 level (2-tailed, and where relevant 1-tailed) was not statistically significant.
Youth who completed the SPARX program were expected to show a decrease in depressive symptoms, which was not evident in this pilot study. However, outcome measures suggest that, for Inuit youth in Nunavut, SPARX may be an effective program for decreasing feelings of hopelessness, as well as the cognitive emotion dysregulation signified by self-blame. Cognitive emotion regulation has consistently been linked to mental health, with those who possess more adaptive regulation strategies better able to cope with stressful or adverse life experiences [31]. Elevated self-blame, rumination, and catastrophizing are 3 factors known to be highly correlated with poor emotion regulation and mental health difficulties, such as depression and anxiety [32]. Youth who tend to ruminate on negative events, catastrophize, and self-blame are more likely to develop depressive symptoms. These youth have increased difficulty inhibiting and regulating their appraisal, which likely leads to more catastrophizing thoughts [33].

### Principal Findings

Overall, key findings suggest that youth who completed the SPARX trial learned new cognitive emotional regulation strategies to help support them in challenging maladaptive thought patterns. Youth also appeared to experience less hopelessness after engaging in SPARX. No formal indication of a decrease in depressive symptoms or an increase in resilience was noted after engaging in the SPARX program.

### Table 2. Overview of outcomes following the SPARX intervention (n=22).

<table>
<thead>
<tr>
<th>Primary outcomes: scale or subscale score</th>
<th>Desirable direction of improvement</th>
<th>Prescore (pooled groups A and B)</th>
<th>Postscore (pooled groups A and B)</th>
<th>Pre to post t test (pooled groups A and B)</th>
<th>Repeated measures ANOVA (pooled groups A and B)</th>
<th>Mixed models (multi-level regression models; comparing groups A and B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYRM(^a): Resilience (\uparrow)</td>
<td>47.389</td>
<td>44.389</td>
<td>Post (↓) ((P=0.02))</td>
<td>Intervention (↓) ((P=0.01))</td>
<td>Intervention (↓) ((P=0.02))</td>
<td></td>
</tr>
<tr>
<td>HSc(^b): Hopelessness (\downarrow)</td>
<td>4.458</td>
<td>3.667</td>
<td>Post (↑) ((P=0.08))^d</td>
<td>Intervention (↓) ((P=0.01))</td>
<td>Intervention (↓) ((P=0.02))^d</td>
<td></td>
</tr>
<tr>
<td>CESDR(^c): Depression (\downarrow)</td>
<td>N/A(^f)</td>
<td>N/A(^f)</td>
<td>Post (↑) ((P=0.02))</td>
<td>Intervention (↑) ((P=0.01))</td>
<td>Intervention (↑) ((P=0.08))</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short(^g) subscale 1: Self-blame</td>
<td>5.348</td>
<td>4.609</td>
<td>Post (↓) ((P=0.01))</td>
<td>Intervention (↓) ((P=0.02))</td>
<td>Intervention (↓) ((P=0.03))^d</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 2: Acceptance (\uparrow)</td>
<td>6.238</td>
<td>6.476</td>
<td>Post (↑) ((P=0.03))</td>
<td>Intervention (↓) ((P=0.07))</td>
<td>Intervention (↓) ((P=0.07))</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 3: Focus on Thought/Rumination (\downarrow)</td>
<td>6.174</td>
<td>5.478</td>
<td>Post (↑) ((P=0.04))^d</td>
<td>Intervention (↓) ((P=0.05))</td>
<td>Intervention (↓) ((P=0.04))^d</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 4: Positive Refocusing (\uparrow)</td>
<td>5.286</td>
<td>5.476</td>
<td>Post (↑) ((P=0.03))</td>
<td>Intervention (↓) ((P=0.08))</td>
<td>Intervention (↑) ((P=0.06))</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 5: Refocus on Planning (\uparrow)</td>
<td>6.905</td>
<td>5.9524</td>
<td>Post (↓) ((P=0.02))</td>
<td>Intervention (↓) ((P=0.02))</td>
<td>Intervention (↓) ((P=0.08, 2-tailed))^d</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 6: Positive Reappraisal (\uparrow)</td>
<td>7.714</td>
<td>6.905</td>
<td>Post (↑) ((P=0.02))</td>
<td>Intervention (↓) ((P=0.07))</td>
<td>Intervention (↓) ((P=0.05, 2-tailed))^d</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 7: Putting Into Perspective (\uparrow)</td>
<td>6.381</td>
<td>5.762</td>
<td>Post (↓) ((P=0.04))</td>
<td>Intervention (↓) ((P=0.07))</td>
<td>Intervention (↓) ((P=0.01))</td>
<td></td>
</tr>
<tr>
<td>CERQ-Short subscale 8: Catastrophizing (\downarrow)</td>
<td>5.435</td>
<td>4.7826</td>
<td>Post (↓) ((P=0.05))^d</td>
<td>Intervention (↓) ((P=0.04))^d</td>
<td>Intervention (↓) ((P=0.03))^d</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)CYRM-12: Child and Youth Resilience Measure (12-item).  
\(^b\)HSC: Hopelessness Scale for Children.  
\(^c\)Approaches significance.  
\(^d\)Significant results \((P<0.05; 1-tailed, unless otherwise indicated).  
\(^e\)CESDR: Centre for Epidemiologic Depression Scale-Revised.  
\(^f\)N/A: not applicable given insufficient data.  
\(^g\)CERQ-Short: Cognitive Emotion Regulation Questionnaire-Short.
Although depressive symptoms on the whole did not decrease after youth trialed SPARX, results suggest a decrease in negative emotion regulation strategies. This is encouraging given the evidence that negative coping styles are associated with greater depressive symptoms [34]. With a reported decrease in self-blame, catastrophizing, and rumination, youth may develop enhanced coping strategies to deal with adverse life events. Hopelessness has been one of the most important mental health factors examined in attempts to understand suicidal behavior and has been linked to suicide in Indigenous populations [35,36]. It is encouraging that results show a decrease in hopelessness and an increase in cognitive emotion regulation strategies. This supports the need to further examine the effectiveness of SPARX on a larger scale as well as more longitudinally, as CBT therapies may be more effective over several months [37].

Youth who completed the SPARX program were expected to experience increased resilience after the intervention but outcome measures provided by youth did not suggest any change in the resilience measure. The small sample size may largely explain this finding. It is also possible that the resilience measure itself was not culturally relevant or appropriate. Despite the lack of significant changes recorded in the formal measures of resilience, there is reason to believe that a decrease in the hopelessness scale may promote and foster greater resilience [36]. It is possible that SPARX may prove to be an effective treatment for decreasing depressive symptoms and boosting resilience when assessed in a larger study.

Youth in Nunavut disproportionately struggle with mental health concerns in comparison to youth in other Canadian territories and provinces. Despite this disparity, Inuit youth lack access to mental health services, largely because of a shortage of staff to provide and assist with evidence-based treatments, a direct result of continued oppression and marginalization under enduring colonialism experienced by these communities [10]. Developing mental health programs that do not necessarily require staff support may be viable options in overcoming this barrier. Extant literature documents the success of computer-based CBT programs in increasing access to mental health resources for youth [38-40]. A review of over 100 studies evaluating the use of e-interventions for children and youth struggling with mental health needs showed that technology-assisted interventions are becoming an increasingly common form of service delivery in Australia, New Zealand, the United States, the United Kingdom, Canada, and the Scandinavian countries [41]. Thus, an increasing number of studies are overwhelmingly supportive of e-interventions, including serious games, as successful therapeutic tools. In addition to their clinical benefit, results suggest enhanced interest in and access to these services by youth, decreased costs compared with traditional services, and greater quality of life for those who access e-interventions [41]. Moreover, a growing field of research is specifically examining the use of these programs in remote and Indigenous communities, with promising results demonstrating the effectiveness of computer-based programs with these populations [38]. Digital interventions have demonstrated success in many remote and underserved communities, as they increase access, break down barriers, and fill key gaps in service provision [41-44].

This study aimed to establish whether SPARX would be an effective intervention for fostering resilience and decreasing the risk of depression with Inuit youth in Northern Canada. It appears as though SPARX may be a good first step for supporting youth with skill development for challenging maladaptive thoughts and providing behavioral management techniques such as deep breathing. Further larger-scale studies are required to understand whether SPARX or similar e-interventions can be effective in decreasing symptoms of depression.

**Limitations**

We encountered several challenges during the implementation of this pilot study: (1) complications with youth recruitment, which resulted in the adoption of less rigid exclusion criteria; (2) challenges pertaining to the staffing of facilitators; (3) unavoidable deadline extensions due to staffing challenges, technological difficulties, and barriers in communication; and (4) a poorly matched control and intervention group when it came to sample size, duration, and frequency of play for youth in each group. The numerous challenges encountered caused this study to be drawn out over 9 months instead of the proposed 4.

As is common with pilot studies, the statistical power of our analysis was limited by a small sample size. Pilot studies are not expected to provide definitive statistical results, but rather to confer suggestive results for follow-up [45], and despite the lack of power we can report some statistically significant findings at the α level of .05. In addition, the “play now play later” procedures for this study were intended to control for the effects of fluctuations in mood due to daylight hours and fluctuating temperatures in the North. This was managed by having youth in the control group exposed to SPARX 2 months after youth in the intervention group completed their use. However, a strict deadline had to be set for the end of the study, and thus many youth in the control group were given a shorter time allowance for completing the SPARX program. Differences between the 2 groups in the frequency and duration of their SPARX play undermined the advantages of having a control group. Further, as a result of this deadline, there were fewer youth who participated in the control group. Finally, for ethical reasons, no participating youth were denied access to SPARX (ie, we ensured both group A and group B participants were all eventually given access to the intervention). Without a true control group where some participants did not receive SPARX, it is impossible to rule out confounds in the results. Unmatched sample sizes across the 2 groups, and an overall smaller-than-anticipated sample size, also weakened the scientific rigor of this study.

Another limitation was that the questionnaires used in this study, which were designed and systematized based on Western, Eurocentric concepts of depression, hopelessness, and resilience, did not properly access the experiences and expressions of mental health for Inuit youth in Canada. It is also possible that challenges with English language literacy confounded findings with Inuit youth participants despite staff being available to assist. It is possible that using local dialects (Inuinnaqtun and...
Inuiktut) for questionnaires would better support youth’s understanding of their mental health difficulties.

Finally, inconsistent access to the internet in most communities meant it was difficult to administer surveys; this also hindered data collection, participation monitoring, and record keeping. The high turnover of community facilitators created confusion with frequent changes in their roles and inadequate time provided for training, knowledge mobilization, and relationship development with youth participants. The conditions under which the trial was administered understandably created ambivalence among community facilitators, who were generally in favor of an intervention that might eventually benefit their communities but also struggled with the additional workload it imposed. All of these barriers, including short staff contracts, high staff turnover, youth attrition, inordinate mental health needs, lack of phone and internet access, and communication difficulties, are closely linked to continued colonization practices. Inuit youth in Canada continue to face systemic challenges and resource disparities as a result of poor government funding and a lack of consultation with communities for needs assessment and program development.

With these barriers and limitations in mind, we are also aware that this pilot study is rooted in a positivistic Western framework and was designed, conducted, and facilitated primarily by Qallunaaq (non-Inuit). However, due to limited funding, this commissioned study had to be conducted remotely and included directives from the Government of Nunavut to involve local mental health workers, most of whom are not Inuk themselves, to administer the SPARX trial. This exploratory pilot served as a feasibility study fostering enthusiasm and cohesion among participating communities and scaffolding more culturally appropriate research. In a companion study [46], participating youth and community facilitators built much of the knowledge necessary through exit focus groups to facilitate such an endeavor. Indeed, efforts are now well underway to follow-up with a more community-directed and culturally meaningful study focused on articulating Inuit knowledge and research epistemologies through holistic, relational perspectives rooted in Piliriqatigiinniq, the Inuit concept for working collaboratively for the common good [20]. We recognize that research in Inuit communities must be collaborative and inclusive, underscoring “the right of colonized, Indigenous peoples to construct knowledge in accordance with the self-determined definitions of what they want to know and how they want to know it” [20](p12).

Future Directions

We present here preliminary findings that suggest that serious games such as SPARX could lead to improvements in emotion regulation and, by association, depressive symptoms; however, conflicting findings suggest youth’s mood did not change after SPARX engagement. It will be important to replicate this study on a larger level, including more communities. It may also be beneficial to allow for online play. This will have the added benefit of increased rigor by allowing the research team to directly monitor engagement and frequency and length of play. It will be important to add a true control condition in the next trial to ensure that SPARX’s effectiveness can be more conclusively established. In addition, it will be important to gauge what mental health and mental illness mean to Inuit youth in Canada. This will help foster an understanding of more culturally appropriate questionnaires for accessing youth’s experiences with SPARX. Finally, using measures that have been validated with Inuit communities in Canada may help ensure that youth can more accurately report their mental health status. With a more culturally appropriate model for understanding Inuit youth mental health symptomology, an assessment of mental health changes attributable to the SPARX program may be more accurate.

Learning from the many barriers encountered here will allow for a more culturally competent and rigorous approach to conducting future research. It will be important to examine these barriers and to address them in any future protocol (such a follow-up study is currently underway). For example, having a community facilitator working specifically on SPARX will better support youth engagement, mitigate attrition, and foster greater trust between the youth and community facilitators. SPARX in schools might synergize well with institutional curricula and help with recruitment, engagement, and the generalizability of skills. Having a researcher on-site in the communities may help with supporting the community facilitator, troubleshooting any technological difficulties, and minimizing communication barriers. These extra supports, and the concomitant potential for more rigorous research with a larger sample size and true control group, will make an effective evaluation of the SPARX program more feasible. Furthermore, it may be advisable to hire a number of youth from each community to promote SPARX, advocate for the program, and help reduce the mental health stigma their communities may be facing.

Finally, it will be imperative to work with youth and communities to design, develop, and test an Inuit version of the SPARX program, tailored to fit the interests of Inuit youth and Elders in Canada and to increase engagement and effectiveness of the program. Such an initiative is underway. Existing studies suggest that incorporating one’s culture into an already compelling psychoeducational serious game has the potential to promote cultural esteem and community-based resilience, in addition to fostering individual resilience.

Acknowledgments

We wish to thank the Nunavut Territorial Department of Health for sponsoring this study. We also gratefully acknowledge youth participants and community facilitators for contributing their time, wisdom, and energy to the project.
Conflicts of Interest

The intellectual property for SPARX is held by UniServices at the University of Auckland. Any proceeds from licensing or selling SPARX outside of New Zealand will be shared in part with UniServices, MS, and MFGL (and other co-developers of SPARX who are not coauthors of this article).

Editorial Notice

This randomized study was only retrospectively registered. The authors explained that this is because it was a small, atypical pilot trial. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered a pilot trial. 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 extension checklist for pilot/feasibility trials.
[PDF File (Adobe PDF File), 155 KB - games_v11i1e38493_app1.pdf ]

References


11. Lindsay T, Healey G. Perspectives of Families Working with Nunavut's Foster Care System. Iqaluit, Canada: Quajigiaartiit Health Research Centre; 2012.


Abbreviations

CBT: cognitive behavioral therapy
CERQ-Short: Cognitive Emotion Regulation Questionnaire
CESD-R: Centre for Epidemiologic Depression Scale-Revised
CYRM-12: Child and Youth Resilience Measure-Short
GNAT: gloomy negative automatic thoughts
HSC: Hopelessness Scale for Children
Co-design of a Virtual Reality Cognitive Remediation Program for Depression (bWell-D) With Patient End Users and Clinicians: Qualitative Interview Study Among Patients and Clinicians

Maria Elena Hernandez Hernandez1, PhD; Erin Michalak1, PhD; Nusrat Choudhury2, MEng; Mark Hewko3, MEng; Ivan Torres1, PhD; Mahesh Menon1, PhD; Raymond W Lam1, MD; Trisha Chakrabarty1, MD

1University of British Columbia, Faculty of Medicine, Department of Psychiatry, Vancouver, BC, Canada
2National Research Council Canada, Medical Devices, Simulation and Digital Health, Montreal, QC, Canada
3National Research Council Canada, Medical Devices, Simulation and Digital Health, Winnipeg, MB, Canada

Corresponding Author:
Maria Elena Hernandez Hernandez, PhD
University of British Columbia
Faculty of Medicine
Department of Psychiatry
2255 Wesbrook Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 822 7314
Email: elena.hernandez@ntnu.no

Abstract

Background: Major depressive disorder (MDD) is the leading cause of global disability; however, the existing treatments do not always address cognitive dysfunction—a core feature of MDD. Immersive virtual reality (VR) has emerged as a promising modality to enhance the real-world effectiveness of cognitive remediation.

Objective: This study aimed to develop the first prototype VR cognitive remediation program for MDD (“bWell-D”). This study gathered qualitative data from end users early in the design process to enhance its efficacy and feasibility in clinical settings.

Methods: Semistructured end-user interviews were conducted remotely (n=15 patients and n=12 clinicians), assessing the participants’ perceptions and goals for a VR cognitive remediation program. Video samples of bWell-D were also shared to obtain feedback regarding the program. The interviews were transcribed, coded, and analyzed via thematic analysis.

Results: End users showed an optimistic outlook toward VR as a treatment modality, and perceived it as a novel approach with the potential of having multiple applications. The participants expressed a need for an engaging VR treatment that included realistic and multisensory settings and activities, as well as customizable features. Some skepticism regarding its effectiveness was also reported, especially when the real-world applications of the practiced skills were not made explicit, as well as some concerns regarding equipment accessibility. A home-based or hybrid (ie, home and clinic) treatment modality was preferred.

Conclusions: Patients and clinicians considered bWell-D interesting, acceptable, and potentially feasible, and provided suggestions to enhance its real-world applicability. The inclusion of end-user feedback is encouraged when developing future VR programs for clinical purposes.

(JMIR Serious Games 2023;11:e43904) doi:10.2196/43904

KEYWORDS
depression; cognitive remediation; cognitive dysfunction; thematic analysis; virtual reality; VR; qualitative study; user-centered design; immersive; co-design; depressive; mental health; mental illness
Introduction

The Socioeconomic Impacts of Depression

Major depressive disorder (MDD) affects approximately 5.4% of the Canadian population [1]. Most recently, the COVID-19 pandemic has increased the proportion of Canadians experiencing depression. Among people aged 25 to 44 years, the proportion of people with MDD increased from 18% in fall 2020 to 23% in spring 2021 [2]. MDD is the leading cause of disability worldwide, substantially affecting psychosocial, occupational, and academic functioning [3]. MDD costs the Canadian economy approximately >CAD $30 billion (US $22 billion) annually in reduced workplace productivity [4]. MDD also affects academic performance in young adults. University students are at higher risk of depression than general population, resulting in lower grade point averages, lower academic performance satisfaction, and higher dropout risk [5-7].

Current treatments, although effective in improving core symptoms of MDD, fall short in addressing MDD-related disability [8-10]. Symptomatic improvements are not necessarily accompanied by restoration of occupational and psychosocial functioning—patients who are diagnosed with mild or remitted depressive symptoms may still show occupational and academic disability [9,10]. Accordingly, treatment priorities have shifted to targeting functional recovery, rather than solely symptomatic remission [8,11,12]. There has been increased interest in expanding mental health services in university campuses, as well as workplace mental health programs to identify and treat employees with depression [13]. While such programs can be helpful in reducing depression symptom severity, they neither necessarily result in major improvements in workplace or academic productivity, nor in decreases in presenteeism [14-18]. Thus, depression treatment strategies specifically targeting academic and occupational recovery remains an unmet need.

Cognitive Deficits in Depression

To enhance occupational functioning in MDD, cognition is an important intermediate target [19]. Cognitive dysfunction is often found in depressive disorders such as MDD, with deficits in memory, attention, processing speed, and executive functioning detectable in acutely symptomatic and remitted patients [20,21]. Importantly, cognitive functioning has been associated with occupational functioning in acutely depressed and recovered individuals with MDD [22-24]. However, there is a lack of effective treatments for cognitive deficits in depression. Major cognitive impairments may persist in individuals who attain symptomatic response or remission with established depression treatments [20], and antidepressant medications have inconsistent and sometimes negative effects on cognitive functioning [25].

Cognitive remediation has emerged as a promising treatment strategy for addressing cognitive deficits in psychiatric disorders [26,27]. This is a treatment method based on the repeated practice of computer or paper-and-pencil exercises that target specific cognitive skills (eg, sustained attention, verbal memory, and working memory) with the purpose of enhancing patients’ work and social functioning abilities [28]. Cognitive remediation has strong evidence for improving cognition and functioning in schizophrenia, as well as showing promise in bipolar disorder and attention-deficit/hyperactivity disorder [26,27,29]. However, these results are less consistent in depression—a recent meta-analysis of cognitive remediation in depressive disorders found moderate effects on attention and working memory, with no major effects on executive functioning or verbal memory [30]. It also remains unclear if the skills practiced during cognitive remediation in depression transfer out of the laboratory environment and result in substantial real-world functional improvements [31-33].

Virtual Reality for the Training of Cognitive Skills

Immersive virtual reality (VR) has attributes which could enhance the delivery of cognitive remediation and has preliminary evidence for efficacy in other neurologically and psychiatric disorders [34]. The immersive and interactive nature of VR facilitates greater user engagement and enjoyment compared with computer presented tasks [35,36]. VR also allows for the practice of skills in naturalistic settings, thereby improving the likelihood of skill transfer to the real world [37-39]. Randomized controlled trials have shown that VR-delivered cognitive and vocational remediation results in substantial improvements in cognition and functioning in patients with traumatic brain injury, stroke, schizophrenia, and substance use disorders, even when compared with active control conditions such as therapist administered remediation [40-43]. Additional advantages of using VR technology—which has been growing in mainstream accessibility and decreasing in expense—include its ability to reach a broader range of individuals with minimal cost. In addition, data collected from VR programs can potentially be integrated into digital health services, allowing a highly personalized intervention.

Despite the burgeoning applications of VR technologies and growing use in psychiatric populations, a sustained course of VR cognitive remediation has not been specifically evaluated in depression. As enthusiasm for using VR in clinical settings has grown, so have the questions regarding how to best design VR clinical applications to be usable and efficacious for the target population. A working group of international experts published the first consensus guidelines for best practices in the design and testing of VR clinical applications [44]. In these guidelines, incorporation of qualitative feedback from multiple end users is identified as a critical component of VR program content development. As lack of end-user involvement in development is often at the root of digital intervention failure [45], following best practices in human-centered design is important to enhance the interventions’ relevance and effectiveness. Feedback from experts and care providers can ensure that the intervention is based on solid theoretical foundations and is feasible for real-world use.

Building on the existing body of work relating to cognitive remediation in clinical populations, the National Research Council Canada and the University of British Columbia have developed the bWell Cognitive Care Platform for Depression (“bWell-D”), a prototype immersive VR cognitive remediation program for individuals with depression. Following best practices in VR clinical application design, this study collected qualitative data from clinicians and care providers regarding
their perceptions and experience with VR, desired outcomes from a cognitive or functional remediation program, and perceived barriers to use of VR cognitive remediation in clinical settings. Later, participants provided their thoughts and opinions specifically regarding the bWell-D program. This information was used to guide program and protocol refinements to bWell-D to boost clinical efficacy and feasibility.

Methods

Ethics Approval
This research was reviewed and approved by the Behavioural Research Ethics Board of the University of British Columbia (Behavioural Research Ethics Board numbers H21-00028 and H20-00746). It was also approved by the Research Ethics Board of the National Research Council Canada (NRC 2020-122).

Participants

Patients
A sample of 15 patients who self-reported a diagnosis of depression were recruited to participate in this study. Recruitment was carried out on the web through the Vancouver Coastal Health Research Institute and REACH BC websites, as well as on a classified advertisements website. Interested participants were screened to ensure they met the eligibility criteria:

1. Aged 18 to 65 years
2. Self-reported diagnosis of previous or current major depressive episodes
3. Self-reported subjective cognitive or functional deficits at baseline, as indicated by a battery of self-report questionnaires. All participants were included in the analysis, regardless of their scores in the clinical measures.
4. Engaged in full-time employment (currently or in the past), with self-reported ongoing functional deficits resulting from depressive symptoms, or off work or on reduced hours with depressive symptoms reported as the primary reason.
5. Sufficient proficiency in English to complete the questionnaires and interview

Patients who were eligible and completed the entire process were compensated for their participation with CAD $20 (US $14.55).

Clinicians

A convenience sample of 12 clinicians participated, with recruitment occurring through the Vancouver Coastal Health Research Institute and REACH BC websites, as well as physical posters displayed on-site at relevant health units. Clinicians were also contacted through the researchers’ professional contacts. Interested clinicians were provided a letter of invitation and confirmed their interest in participating via email. Clinicians must have been allied health professionals (eg, psychiatrists, psychologists, family physicians, occupational therapists, or nurses) with experience working with populations with depression, and experience with cognitive or functional remediation in psychiatric populations. Clinicians were not offered a monetary compensation.

Procedure

This study was conducted remotely through videoconferencing. Participants attended an approximately 1-hour Zoom interview, in which they were first explained the purposes of the study, and oral informed consent were obtained. Participants then completed a web-based Qualtrics survey asking for demographic data. In addition, patients completed a series of psychiatric and psychological questionnaires (see the Measures section), whereas clinicians answered questions related to their clinical practice. After completing the questionnaires, a semistructured interview was conducted, in which participants were asked regarding their perceptions and experiences with VR, their desired outcomes from a workplace cognitive or functional remediation program (bWell-D), and perceived barriers to use this type of treatment in a clinical setting. Participants were then shown previously recorded video samples of the bWell-D tasks and were asked to provide their thoughts on them, their potential relevance to daily life, and suggestions on how to improve the tasks to enhance engagement and real-world applicability (see Multimedia Appendix 1 for the interview scripts).

The bWell-D Program

Overview
“bWell-D” is a prototype of immersive VR cognitive remediation program for depression, developed by a collaboration between the University of British Columbia and the National Research Council Canada. The foundational platform, bWell, was developed within a network of researchers and clinicians across Canada, features configurable exercises and design as a broadly applicable toolkit targeting general aspects of cognition commonly affected across many disorders [46]. Preliminary acceptability studies in healthy individuals indicate that the platform is enjoyable, engaging, and well tolerated [35,47].

bWell-D is a customized version of bWell, targeting cognitive and real-world challenges germane to MDD depression [20,21,46]. Two bWell tasks were modified to include real-world environments, that is, the (1) office and (2) classroom. It additionally uses previously identified common components of successful cognitive remediation in psychiatric populations. These include (1) "errorless learning," wherein participants are provided support in learning all task components, with supports gradually removed as skills and confidence increase and (2) “adaptivity,” in which tasks become increasingly complex to match with the participant’s competence [31]. bWell-D involves 4 different tasks (Figure 1) that aim to train cognitive skills that are commonly affected in MDD. Participants can interact with the virtual environment through a headset and 2 hand controllers.
Egg (Attention)
In this task, the patient is asked to look around in an office environment and look for eggs. Once locating the egg with their sight, they are required to fix their gaze on the egg until it hatches.

Lab (Multitasking)
Participants must complete 2 recipes simultaneously by pouring multicolored flasks into 2 mixing bowls. The recipes will appear in 2 tablets placed nearby, and participants will have to go back and forth between recipes.

Mole (Reaction Time and Response Inhibition)
Following similar rules to the whack-a-mole game, participants hold a hammer in each hand and hit the cylinders that pop out of a table. The cylinders are multicolored, and the colors of the hammers also change over time. Participants are asked to hit the cylinder that matches the hammer’s color.

Theater (Visual Working Memory)
The participants are shown a sequence of shapes that will be hidden after a set viewing time. A pool of objects will then be presented to the participant, from which they must choose the shapes that were initially shown to them and recreate the original sequence.

The bWell-D tasks operate in three modes: (1) tutorial, (2) assessment, and (3) training. Tutorial has the aim of familiarizing the patient with the virtual environment and showing them the actions they must perform in the tasks.

Assessment mode has the purpose of identifying the patients’ individual needs by assessing their task performance with a fixed difficulty level. In the training mode, the level of difficulty is adaptable based on the patients’ performance. Real-time feedback is provided to the patient, showing them level or score changes and successes and errors.

Measures
Along with the demographic questions included in the Qualtrics questionnaire, patients also responded to the following psychiatric and psychological measures:

Lam Employment Absence and Productivity Scale
The Lam Employment Absence Productivity Scale [48] is a work functioning and productivity scale to be used specifically with patients with depression. It is a 10-item scale, in which the first 3 questions are open-ended and ask regarding the patients’ type of work, the number of working hours scheduled in the past 2 weeks, and the number of working hours missed in those 2 weeks. The following items are rated on a 5-point Likert scale (“None of the time” to “All the time”), scored as 0 to 4. Work impairment rating can range between “None to minimal” to “Very severe.”

Generalized Anxiety Disorder–7 Items
The Generalized Anxiety Disorder–7 items [49] is a short measure of anxiety. Participants are asked how often they have been bothered by their anxiety symptoms during the past 2 weeks. The measure follows a 5-point Likert format, ranging from “not at all” (0) to “nearly every day” (4).
Sheehan Disability Scale

The Sheehan Disability Scale [50] is a short measure of disability and functional impairment. The first 3 questions are in a Likert scale format, inquiring regarding the severity of cognitive symptoms affecting the patients’ work and social and family life from “not at all” (0) to “extremely” (10). The last 2 open-ended questions ask regarding the patients’ missed days at work, and regarding the days when the patient felt underproductive.

Perceived Deficits Questionnaire–Depression

The Perceived Deficits Questionnaire–Depression [51] is a 20-item questionnaire that assesses subjective cognitive dysfunction in people with depression in the past 7 days. It follows a 5-point Likert scale, with response options ranging from “Never” (0) to “Almost always” (4).

Beck Depression Inventory

The Beck Depression Inventory [52] is a 21-question, self-report Likert scale that measures the severity of depression symptoms. Responses range from 0 to 4, in which a higher number indicates a more pronounced symptom severity.

Cognitive Failures Questionnaire

The Cognitive Failures Questionnaire [53] is a measure of self-reported failures in perception, memory, and motor function over the past 6 months. It is a Likert scale consisting of 25 items ranging from “Never” (0) to “Very often” (4).

Qualitative Data Analysis

Interviews were recorded and transcribed verbatim. The transcriptions were made with Otter.ai, a web-based service that uses artificial intelligence to transcribe audio recordings. The transcriptions were reviewed and corrected by hand after being processed by Otter.ai. The initial 5 interviews (3 clinician interviews and 2 patient interviews) were coded by the first author to identify an emerging thematic framework [54]. Coding was conducted separately for the clinician and patient groups. However, members of the research team (EM, TC, and MEHH) discussed and identified emerging patterns in the data and determined that the coding framework between the 2 groups was similar; therefore, it was decided to analyze the 2 groups together. Subsequent transcripts were coanalyzed and discussed based on the initial thematic framework. Thus, the data analysis began with the delivery of the first interview and proceeded concurrently and iteratively. The NVivo (QSR International) software was used to aid with the coding and thematic analysis process.

Results

Participant Characteristics

Patients

Participants were aged between 21 and 63 years, with a median age of 31 (mean 34.2, SD 11.5) years. Most participants self-identified as women (n=10), with 3 as men and 2 as nonbinary. The most common diagnosis was MDD (n=8), and 1 patient was diagnosed with dysthymia, 1 with bipolar disorder, 4 of them reported being diagnosed with depression but did not remember the specific diagnosis, and 1 did not have a formal depression diagnosis, but manifested having all the clinical symptoms for it (this was later confirmed at the screening stage). Patients with a formal diagnosis reported being diagnosed between 3 and 16 years ago. At the time of the interview, none of the patient participants were undertaking any psychological treatment, however, 4 of them were taking medication to treat their depression symptoms.

Patients’ clinical scores are provided in Table 1. Overall, participants surpassed the threshold for a clinical diagnosis in all measures. All participants reported having their work or academic performance affected by their depression symptoms, and all of them have had to take some time off work or school, ranging from a few sporadic days, to several years. Participants stated working only approximately 80% of their scheduled work hours, some of them having to quit work completely because of the severity of their depression symptoms. Furthermore, 9 patients were either currently employed or studying, while 6 of them were unemployed and identified depression as one of the main causes for their unemployment. From the participants who were employed at the time of the interview, 4 had administrative jobs, 3 had professional jobs, and 2 worked in sales.
Table 1. Patients’ clinical measures.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Values, n</th>
<th>Values, median (range)</th>
<th>Severitya</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEAPSb,c</td>
<td>9</td>
<td>15 (4-29)</td>
<td>Moderate</td>
</tr>
<tr>
<td>GAD-7d</td>
<td>15</td>
<td>19 (9-27)</td>
<td>Severe</td>
</tr>
<tr>
<td>SDS e</td>
<td>15</td>
<td>22 (6-30)</td>
<td>Markedly impaired</td>
</tr>
<tr>
<td>PDQ-Df</td>
<td>15</td>
<td>46 (7-69)</td>
<td>Risk range</td>
</tr>
<tr>
<td>BDPg</td>
<td>15</td>
<td>25 (8-47)</td>
<td>Moderate depression</td>
</tr>
<tr>
<td>CFQh</td>
<td>15</td>
<td>62 (15-79)</td>
<td>Above average cognitive complaints</td>
</tr>
</tbody>
</table>

aOn the basis of median scores.
bThe LEAPS is only for participants who are currently engaged in paid employment, hence the reduced sample who responded to this specific measure.
cLEAPS: Lam Employment Absence and Productivity Scale.
dGAD-7: Generalized Anxiety Disorder–7 items.
eSDS: Sheehan Disability Scale.
fPDQ-D: Perceived Deficits Questionnaire–Depression.
gBDP: Beck Depression Inventory.
hCFQ: Cognitive Failures Questionnaire.

According to the open-ended questions in the Lam Employment Absence Productivity Scale, employed participants had between 50 and 80 scheduled work hours in the previous 2 weeks, and although most of them worked their full shifts, 2 participants only worked between 15 and 20 hours. Four participants were currently working reduced hours, and they have been in this work modality between 2 months and 1 year. The responses to the open-ended questions of the Sheehan Disability Scale indicated that employed participants had lost between 1 and 20 days of work in the past month because of their depression symptoms. Even when attending work or school, participants reported that depression affected their performance in the last month, ranging from 30% of the days to 50% of the days. One participant reported feeling these effects daily.

Clinicians

Participants self-identified as 5 women and 7 men. There were 7 psychiatrists, 2 occupational therapists, 1 psychologist, 1 clinical counselor, and 1 neuropsychologist. They had varied years of experience as mental health practitioners, ranging from 2 to 50 years, with a median of 17.5 years. Five clinicians reported having <5 years of experience in clinical practice, while 6 of them reported having >20. One clinician reported having between 5 and 10 years of experience.

Most of the clinicians worked in an outpatient setting (n=10), either at hospital clinics or private practice. Only 1 clinician worked at an inpatient psychiatric hospital, and another one worked at a community health center. In total, 7 clinicians provided a consultation and assessment and short-term follow-up model of care, 3 of them offered long-term follow-up, and 2 of them did single consultation and assessment with no follow-up. Furthermore, 4 clinicians offered primarily pharmacological and medication treatment modality, 3 of them offered primarily psychotherapy, 3 of them a combination of medication and psychotherapy, and 2 focused primarily in functional rehabilitation. Clinicians who provided psychotherapy offered behavioral treatments, such as cognitive behavioral therapy, dialectical behavior therapy, and acceptance and commitment therapy.

A total of 7 clinicians estimated <50% of their patients had a depressive disorder, while the remaining 5 estimated that more of the 50% of their patients had depression. Five clinicians had experience with cognitive or functional remediation, and they had between 0.5 and 5 years of experience with this type of treatment. Most clinicians (n=9) had been working remotely because the start of the COVID-19 pandemic, mostly relying on videocalls, phone calls, and (less commonly) computer tasks with their patients.

Thematic Analysis

Overview

Participants held mostly positive opinions regarding bWell-D, considering it an entertaining and safe treatment modality. Five themes were identified during the thematic analysis, specifically participants’ previous knowledge regarding VR technology, its potential for clinical uses, a desire for realistic settings and situations, the entertaining features of VR, and the need for an inclusive treatment. A description of each theme and representative quotes are presented below.

Familiarity and Optimism: Openness to Trying VR for Mental Health Symptoms

Both patients and clinicians had at least some degree of familiarity with VR. Patients were familiar with VR mostly from videogames and movies, while clinicians were also aware of its use for clinical purposes, especially for anxiety-related conditions, psychosis, and brain injury. When participants did not have direct experience with VR, they mentioned hearing good things regarding it. Participants with little or no exposure to the technology spoke of their curiosity regarding trying it.
Participants with a higher exposure to VR, and the ones who enjoyed videogames, were the ones who showed higher enthusiasm regarding this technology.

**VR as a Potential but Underexplored Treatment Option for Cognitive and Functional Deficits in Depression**

Although both patients and clinicians often considered mood symptoms as the most salient symptom in depression, they acknowledged that cognitive symptoms are also affected. They described cognitive issues as particularly harmful for work life, because they have a direct effect on job performance, which can in turn exacerbate feelings of shame and stigma:

> It was just a really hard time. I just, I couldn’t concentrate, well, I still have a hard time concentrating. Just my, my work ethics went down. I don’t know, everything became really hard. I missed a few days here and there, like I felt, but at the same time, like even though I was told by the doctors like, ‘No, you need to take time off,’ I kept going because I didn’t want to be judged behind my back. But if anything that made it worse [...] But I know that my, my work has gone down, like, and I have a really hard time concentrating. [Patient 001; woman; 21 years]

Several patients considered that training their cognitive skills through VR can help them not only for work purposes but also with their day-to-day life:

> I go into a room to get band aids, I forget what I’m in there for, literally my house is not that big. And I go back in the living room wondering what the hell I was looking for. So it, I feel like it would help with work and with my personal life, which is just a double bonus. [Patient 003; woman; 43 years]

They were confident that their cognitive abilities can be re-established after experiencing depression. When asked regarding the potential applications of VR for depression, 2 clinicians and half of the patient sample struggled to see how this technology could be used for this condition:

> I’m aware of the promise of virtual reality treatments for anxiety disorders and things like that. I would be, I don’t see the application to depression or bipolar disorder. So you know, in that end, those are the two disorders I treat the most. So therefore, I’m not particularly looking at this. And I’m surprised to hear about a virtual reality treatment for depression [Clinician 009; man; psychiatrist]

Participants considered VR a safe alternative to conventional psychiatric treatments, and most patients were willing to try it because of its perceived lack of serious side effects. Participants believed that more effective treatments for mood disorders are needed, and therefore, VR can be a promising alternative.

> I don’t always enjoy the medication approach, like the side effects can be quite frustrating at times. So I think something without, like, a less medicalized approach could be really helpful. And you could avoid some of the negative side effects. [Patient 014; woman; 21 years]

> I think we need all the treatments we can get for mood disorders. So I’m curious. [Clinician 009; man; psychiatrist]

A few participants, however, considered this “newness” a limitation to start using VR, because it is not standard practice in most health care institutions:

> I practice in a large university hospital type setting. The university hospital always is incredibly bureaucratic, and never wants anything that’s innovative or different. They only want things that are standard. So anytime you try to introduce something new, it’s very challenging. So I would say that it would be unlikely that even if this treatment were approved tomorrow by Health Canada or the FDA, I would say it’s very unlikely that our system would use it for anytime in the next few years. [Clinician 009; man; psychiatrist]

Despite the general enthusiasm regarding using VR and cognitive remediation, both patients and clinicians had some further reservations, and mentioned needing more evidence before undertaking a treatment of this nature. Most clinicians mentioned that, although this type of treatment has shown promising results in some conditions, they would like to have more evidence for its specific use in depression. They would like to obtain this information from methodologically sound research studies:

> If there is going to be evidence for that, for the type of patients that I’ve seen, and the patients are willing to participate? So yeah, for sure [...] It’s not part of the kind of the regular practice, or part of the standard guidelines. But once the evidence is there, for sure. [Clinician 012; man; psychiatrist]

Furthermore, approximately half of the patient sample mentioned that the VR format might not have any particular advantages over other modalities of cognitive remediation delivery:
I think the only like, pushback that might be received is like, what is the difference between doing this and playing like, a task related like, video game, I guess? Like, I think that there’s an, there’s maybe an opportunity for, like, people who are depressed to be like, “Well, if I if I’m basically just going to a lab to play a game, what’s the difference between me doing that, and me just like downloading a game on my phone where you’re being asked to, like, do memory parts?” [Patient 008; woman; 30 years]

**Desire for Ecologic Validity**

Several patients mentioned that they would like to understand exactly how the tasks practiced in bWell-D work, so they do not feel like they are “just playing a game.” They also mentioned needing an explanation regarding how to extrapolate the trained skills to their real-life world:

My honest opinion on the games, they all seem like kid games. They all seem very simple and straightforward, but simple animation. And I find like, I might think it’s just a silly game, but I think once I start actually playing them, I would realize how it’s helping. [Patient 007; woman; 28 years]

I think I would need the therapist to, like, fully explain what this was training and helping with, because, like, the one thing that triggered me when I was in a depressive state was that, like, I knew I was smart, but then like, when I was in a depression, I wasn’t functioning, that, like, as normal, and then I would feel really, really dumb. And like, seeing this kind of activity makes me feel like, you know, it gives a reaction like, I’m not a child, like, why am I putting shapes together? [Patient 10; woman; 27 years]

Most participants mentioned that VR offers the possibility of immersing themselves in settings and situations that resemble real life in a highly accurate manner:

It’s very interesting in the way that all new technology is very interesting. And it’s remarkable, you know, the sense of presence when you’re in the VR environment. It, you know, unless you’ve experienced it, it’s hard to imagine, you know, it’s very much like being in the real world. [Clinician 011; man; psychiatrist]

Most patients mentioned they would like the virtual environments to look like their own work or school setting, as well as including objects (eg, office supplies or school furniture) and situations (eg, conversations with colleagues, following instructions given by their boss or teacher) that emulate real life (eg, interactions with humans and distractions). They believed that these naturalistic aspects can enhance skill extrapolation:

I think that it would also require some effort relevant to their daily activities, or have some translational component. [Clinician 004; man; psychiatrist]

I think if I work in an office, it would, it would be better because I’d be able to take what I’m doing within that virtual reality, and apply it more to the environment that I’m hoping to improve my concentration in. But because it doesn’t look like where I work it’s difficult to connect the two. [Patient 007; woman; 27 years]

Participants considered VR as “the second-best thing to real life,” because it allows an immersion in highly realistic environments. Participants desired to obtain this realism through several ways, (1) by having virtual settings that emulated patients’ own real-life environments, (2) by including distractors in the VR world that would occur in real life, and (to a lesser extent) (3) by having realistic-looking graphics. Patients and clinicians also stated that VR environments allow for the practice of skills and exercises in a more controlled and less intimidating setting:

If I were able, was able to get better at that, and because it’s a low stress environment in the game, maybe having it, having that experience in a low stress situation would help me in my more stressful situations. [Patient 014; woman; 21 years]

It’s like the traditional thing when you, when you’re nervous about speaking in front of a crowd is that, you like, you practice your speech in the mirror. It could be something similar to that but in a VR setting, with there being more of a...I don’t quite know how to articulate it, exactly. [Patient 005; nonbinary; 25 years]

A few patient participants mentioned that VR can allow them to focus on only training their skills without getting sidetracked. However, they emphasized the need for improving their cognitive symptoms overall, rather than just improving their performance in the virtual tasks:

I’m skeptical because of the literature that’s behind this, because there’s a lot of evidence saying that you only get better at the game instead of the skill itself, or the cognitive aspect that you’re focusing on itself. [Patient 001; woman; 21 years]

Three clinicians mentioned the importance of ensuring that the skills practiced in programs like bWell-D are actually training what they intend to train, and that the tasks are representative of the actual cognitive processes that patients must perform at their work or day-to-day life:

If there’s evidence that depressed people have trouble on that sort of task, it makes sense. If you’re pulling it out of thin air, because they have trouble on an attempt and inhibition task that’s not like that one, then you better rethink how close that task mimics the process you’re interested in. [Clinician 011; man; psychiatrist]

Approximately a quarter of the participants were concerned regarding the long-term effectiveness of the treatment. To address this, a patient suggested having a longer VR cognitive remediation treatment, whereas a clinician suggested adding booster sessions. Several participants expressed a desire to obtain treatment progress data through a reliable instrument. Although a few of participants mentioned wanting improved graphics in bWell-D (more realistic looking), others mentioned that this is not too important if the content of the tasks is well
designed, and the objects are clear to see. Most participants mentioned that any possible glitches and malfunctions should be minimal or nonexistent for the treatment to be effective and enjoyable.

**A Fun and Engaging Treatment: The Advantages of a Digital Environment**

Most participants mentioned that the gamified experience of VR adds fun, engaging, and immersive elements to the treatment. A need for multisensory stimuli (i.e., tasks that incorporated auditory and visual components) was mentioned by most of the participants and was considered key for the immersion and involvement in the virtual tasks. Participants also stressed the importance of having a difficulty progression in the tasks, to enhance the challenging aspect of the treatment (without it turning so difficult that it becomes discouraging). A clinician mentioned how this difficulty progression can provide a sense of accomplishment for patients:

*It definitely gives you kind of that drive, while at the same time being complicated and interesting enough to keep your attention wanting to keep going [...] – I found that one was always a really fun task, a difficult task. Not difficult, but it required enough energy without being overwhelming.* [Patient 015; woman; 33 years]

Most clinicians and patients stressed the importance of having a simple, user-friendly VR interface that included some training period and clear instructions to practice the tasks. A few clinicians considered important to assess patients’ satisfaction with the program session to session.

Several participants, especially clinicians, found the versatility of VR to be beneficial for the treatment. They noted that the flexibility provided by VR allows a wide variety of virtual settings, going beyond what is achievable in real life:

*I think you can help us create a very realistic environment to do some therapeutic work in and potentially make some progress that we haven’t been able to outside.* [Clinician 007; woman; psychologist]

A few participants considered a treatment, such as bWell-D, less emotionally draining treatment modality than medications or talking therapy:

*When I was watching it, it occurred to me it’s kind of nice to have these videos not relating to real life experiences, and things that are supposed to be kind of fun, because it doesn’t feel like you’re doing work, even though you are. And that is, at least for me, isn’t as emotionally or cognitively draining.* [Patient 015; woman; 33 years]

Seven participants (both patients and clinicians) mentioned that the game-like setting of VR also allows for progress tracking (e.g., score and level changes and improvements), either from the beginning to the end of a session, or from session to session.

**One Size Does Not Fit All: The Need for Inclusivity and Customizability**

Most patients and clinicians mentioned that they would like to have some agency in the manner the VR treatment is conducted. One of the most recurring comments among patients and clinicians was a preference for the patient to have the treatment individually at home, either partially or completely. Participants believe this modality provides a sense of comfort and control:

*It seems like a nice way to do it yourself. Like, in your own time in, in quiet, in a quiet environment and not have to go anywhere.* [Patient 002; woman; 36 years]

*I think it would be helpful for people who can’t get there, like remotely. Like, there are days when I can’t like, get out of bed or like, get overwhelmed by taking the bus, and like, that might be super helpful to be able to join in like that.* [Patient 10; woman; 27 years]

However, a few patients and clinicians believed a treatment of this nature would be better performed at a clinic. In particular, patients with limited space, or with suboptimal conditions to carry out the treatment at home, would like to have the option to do it at a clinic if necessary:

*If I had to use it at my own home, that would be almost impossible with three kids and, I think, life.* [Patient 012; woman; 42 years]

Participants mentioned wanting a person involved in the treatment even if they choose to do it remotely. Several patients stated that the involved person should be a well-trained facilitator who they trusted. Most of the patients and clinicians mentioned they would like to engage in a VR treatment when mood symptoms and other potential comorbidities are under control, and therefore had enough motivation to undertake the treatment:

*It depends on the severity, so I think it could be quite hard for some individuals to sustain, you know, attention or memory, if they’re having more severe symptoms. But I think as a symptom, sort of subside or set, you know, it’s more mild to moderate, we might be able to focus a bit more and pay attention to it a bit better.* [Clinician 007; woman; psychologist]

*I think like when you are depressed, like, it’d be really hard [...] if there’s any sort of resistance in that process. I think like, a really anxious person or really depressed person would be like, ‘Nope, I can’t deal with it. It’s not working. I’m done with it. Not going to use it today’ sort of mentality.* [Patient 008; woman; 30 years]

In fact, several patients and a few clinicians mentioned the usefulness of including an activity in bWell-D that specifically targeted mood. After looking at the task video samples, 4 patients also mentioned that some tasks could be more useful than others (e.g., patient not currently struggling with memory issues or wanting to focus more on multitasking). Therefore, they would like to tailor the treatment according to their own cognitive struggles and needs. A clinician also commented that, although the office or school-like VR settings could be useful
to emulate naturalistic settings, these settings can trigger emotional responses from some people:

For people who are not working, many people with depression are on short or long term disability from work. I was just wondering if seeing the office environment might trigger something. [Clinician 010; woman; psychologist]

Despite the overall positive perceptions regarding bWell-D, approximately half of the clinician sample and one patient stated they would like to do a VR treatment in parallel with other conventional depression treatments (eg, medications or psychotherapy):

You target that symptom, but then you work with the clinician long term to, to use this as sort of one of the other interventions that you’re using, and I would see this not being a single intervention being used. I see this being used in combination. [Patient 001; woman, 21 years]

Another of the most mentioned topics among patients and clinicians was how age could be an important factor when undertaking a VR cognitive remediation treatment. In particular, participants believed that the more technologically savvy individuals (often the younger generations) would like the treatment better. In turn, participants believed that older patients might need more coaching and adjustment time to become used to bWell-D. Several participants also believed that people prone to sensorial issues might dislike a VR treatment, because it can make them feel physically ill:

The other problem with all of this is that people go into their depression with some of them, depression of course worsens all of this, it can worsen this, but some people have pre-existing audio processing difficulties or visual processing difficulties either genetically or by head injury or a stroke or medical injuries, etc. [Clinician 002; man; psychiatrist]

Participants also mentioned that the treatment should be accommodating to people from different backgrounds and with different needs, such as people with color blindness or visual impairments, people whose English is their second language, and people with limited mobility. One of the most common concerns was the actual access to the bWell-D program and the necessary equipment (eg, headset, controllers, or computer). Most participants acknowledged that these technological devices, although they are becoming cheaper with time, may still be inaccessible for people with more limited economical resources:

They [VR sets] cost a little bit of money, and most people with mental illness don’t have any money. So yeah, just having access to it, having access to a headset [Patient 004; man; 49 years]

The job that I work, I don’t have coverage yet, and when I do get coverage, I don’t know how much it is. And maybe, I’m not sure if the city offers it, like if it’s readily available. So I guess cost and availability would be the biggest challenges for that. [Patient 007; woman; 28 years]

However, a clinician mentioned that, if the treatment shows to be effective for patients, then every effort should be made to open access to it.

Changes Made to bWell-D After Qualitative Feedback

Following the guidelines proposed by Birkhead et al [44], the changes suggested by end users to improve the bWell-D program were discussed among the research team, translated into representative software functionalities during team brainstorming sessions, and were given priority for implementation based on (1) how frequently these changes were suggested by end users and (2) whether these changes were feasible for development in a VR environment. The main areas of improvement suggested by both clinicians and patients were based on the thematic analysis, and mainly centered in (1) taking greater advantage of the potential of VR to create more multisensorial tasks; (2) the inclusion of more ecologically valid VR elements (to reflect more realistic VR environments), as well as having bridging exercises to relate the skills trained in bWell-D to real-world situations; and (3) providing sufficient challenge for fun and engaging tasks. In addition, end users suggested having a user-friendly interface and tutorials to have an opportunity to practice the tasks, which they considered especially important for an at-home treatment. Moreover, participants recommended to include an additional VR activity that specifically targets their mood. Participants believed that these modifications could make bWell-D more interesting, engaging and useful to improve cognitive deficits. The implemented functionalities are described in Multimedia Appendix 2, while Multimedia Appendix 3 provides a more in-depth description of these changes, and Multimedia Appendix 4 shows some visual examples.

Although efforts were made to cover as many of the suggested changes as possible to bWell-D, there were some areas that the research team determined to address with alternative strategies. In particular, one of the suggestions made by several participants was the addition of more VR environments that better emulated patients’ real-life settings, such as their own workplace or school. However, the research team concluded that the addition of these VR environments would be too demanding in terms of resources required for software development. Instead, it was decided that such need would be addressed through bridging exercises, which are clinician-led discussions that help cognitive remediation participants to apply what is trained in the tasks to everyday life [55].

As part of the iterative process in the design of bWell-D, the changes made to bWell-D were later shown to a subset of 4 patient participants, once more as prerecorded video samples. After looking at the updated version of the program, participants were provided with additional information regarding bWell-D and were asked regarding their general opinions regarding the implemented changes (see Multimedia Appendix 5 for the reinterview script). All the participants reacted positively to the changes, and considered them an improvement to the initial version of the program.
Discussion

Principal Findings

The main aim of this study was to explore the opinions of patients and clinicians regarding the use of VR for the treatment of depression-related cognitive dysfunction. bWell-D, a VR treatment specifically designed for treating cognitive symptoms of depression, was used as the main VR treatment example for the participants. The results indicate that VR interventions for treating depression and its cognitive symptoms are generally accepted by both patients and clinicians. In the particular case of bWell-D, they considered it an innovative and fun alternative for the treatment of mental health issues. Some suggestions were made for the improvement of bWell-D, such as including multisensorial tasks, cues, and distractors, a progression in the difficulty of the tasks, realistic virtual settings and activities, having a user-friendly interface, including tutorials and an opportunity to practice the tasks before initiating the treatment, and adding a task or VR activity that targets patients’ emotional state.

Although the general perceptions regarding VR for clinical purposes and bWell-D were mostly positive, participants also expressed some reservations. In particular, clinicians needed more evidence before implementing a treatment of this nature in their own clinical practice. Although there is a growing body of evidence supporting cognitive remediation [56] and VR in health care settings [57], structural obstacles within health institutions, as well as health professionals' own personal preferences (ie, an inclination for more traditional therapeutic approaches), can prevent them from being open to use strategies or technologies that deviate from their usual practice. This phenomenon has been previously described as resistance to change [58], and it is one of the main reasons why novel approaches in the treatment of several conditions remain underused [59].

Similarly, patients mentioned wanting a thorough explanation regarding how the skills trained in programs such as bWell-D can extrapolate to the real world. To this end, psychoeducation has been identified as a key component in the treatment of depression [60], which could address the utility of training cognitive skills as a means to achieve recovery. This psychoeducation component has been found useful by patients in previous studies, such as the study by Lindner et al [61]. It is worth mentioning that 3 of the younger participants considered the VR tasks too infantile or simple to relieve cognitive symptoms of depression, which could be an indication of a generational preference regarding the type of tasks offered through bWell-D. As a result, the inclusion of realistic paraphernalia in the VR environments, as well as the increasing difficulty of the tasks, can potentially address these concerns.

Patients also commented regarding the need of a customizable VR treatment in which they could decide over how, when, and where the treatment was conducted. There was a strong preference for doing the treatment individually at home, because this was considered the most convenient modality for patients. Although considering patients’ opinions within any psychiatric or psychological intervention is essential, clinicians should be mindful regarding not losing key components of the intervention when accommodating patients’ preferences. By doing so, clinicians might be drifting away from best practice, preventing patients of receiving the best treatment available [62].

There was an overall positive perception toward VR interventions for depression and other mental health conditions, and toward bWell-D itself. Participants indicated that they would use this type of treatment if available, and this acceptance was mainly motivated by the fact that a VR intervention was perceived as a treatment with no adverse long-lasting consequences or side effects. Participants believed that, in the worst-case scenario, the treatment may not improve their cognitive skills but would not worsen them, which was an advantage when compared with other more conventional treatments, such as medication. This relative safety provided by VR treatments can be highly beneficial by end users, given that, besides the economical investment, there are little to no risks associated with this treatment. However, the accessibility concerns manifested by end users should be a consideration for implementation and future directions in the bWell-D development process, given that currently there are no clear plans regarding how this treatment will be made available for clinicians and patients.

This study shares some specific findings with previous literature. In the study by Kramer et al [63], clinicians also indicated a need for more empirical evidence to support the use of VR as an assessment or therapeutic tool, and they also expressed a concern regarding how the use of VR faces challenges regarding its adoption and assimilation. In the study by Thompson et al [64], participants were also worried regarding technological difficulties, mentioning how they might not have access to the necessary equipment to undertake the treatment, and how including a VR intervention in their routine might interfere with patients’ schedules. The lack of perceived adverse effects was also highlighted in the study by Nason et al [65], in which patients experienced only minimal motion sickness. As in our own study, Krebs et al [66] found that relatable humanoid avatars are important in virtual environments, given that they promote the suspension of disbelief, which is the key difference between a gamified intervention and a more traditional one. The results of our research join the large body of studies in which VR interventions for mental health issues are perceived positively by end users [67-69]. However, to the best of our knowledge, this is the first study that analyses qualitatively the use of cognitive remediation for cognitive symptoms of depression.

Limitations and Strengths

This research had some limitations. Further insights might have been obtained from participants if they have had hands-on experience with bWell-D, instead of watching prerecorded samples. While it was initially intended that participants would trial bWell-D before completing the qualitative interview, COVID-19 pandemic restrictions prohibited on-site testing during the data collection period, necessitating the change to the virtual format used here. Related to these COVID-19 pandemic restrictions, patient participants were recruited through various web-based sources and self-reported their depression.
symptoms. Future studies might benefit from in-person assessments in which a health care professional administers the clinical questionnaires. In addition, most of the clinician sample did not have practical experience with cognitive remediation, clinicians with direct experience could have clearer expectations of a VR cognitive remediation treatment and may have therefore provided more extensive feedback. The sample included both people who were still working or in school but had depression-related impairments, as well as people who were off work—these 2 groups might conceivably have different concerns and needs from cognitive remediation. Finally, the clinician sample was formed mostly by psychiatrists working in outpatient settings—a more diverse clinician sample might have provided broader perspectives regarding the use of cognitive remediation and VR for depression. In terms of strengths, this is, to our knowledge, the first study that focused on obtaining end-user feedback regarding a VR intervention for cognitive symptoms of depression with a qualitative approach. We were able to obtain rich and extensive feedback from a varied sample that included patients from a wide range of ages, backgrounds, and genders, which offers an inclusive and comprehensive perspective toward the use of VR and cognitive remediation.

Finally, reinterviewing the participants reaffirmed the relevance of the changes made to the program and enhanced the user-centered design, which will continue to be central in the iterative process of bWell-D development until its application in clinical populations.

Conclusions

Virtual interventions for the improvement of cognitive symptoms in depression are generally acceptable and potentially feasible for patients and clinicians. Although VR and its use for the treatment of health-related conditions is relatively new, participants were familiar with it and open to try it, given its perceived safety. They were curious regarding a treatment of this nature and saw potential in its implementation. Participants considered it to be important to have realistic virtual settings, tasks, and cues to enhance the extrapolability of the skills trained in the VR program. The virtual format of this intervention was perceived as an advantage, and participants believed it offered a versatile and fun alternative to conventional depression treatments. Patients’ own needs and preferences should be considered when implementing a VR treatment for psychiatric conditions, and all efforts should be made to make it more accessible and convenient for end users.

Conflicts of Interest

IT has received research funding from the Michael Smith Health Research and BC Schizophrenia Society Foundation and Canadian Institutes of Health Research, and has served as consultant for Lundbeck Canada, Sumitomo Dainippon, Boehringer Ingelheim, and Community Living British Columbia. RWL has received honoraria for ad hoc speaking or advising and consulting, or received research funds, from Abbvie, Asia-Pacific Economic Cooperation, Bausch, BC Leading Edge Foundation, Brain Canada, Canadian Institutes of Health Research, Canadian Network for Mood and Anxiety Treatments, CAN-BIND Solutions, Carnot, Grand Challenges Canada, Healthy Minds Canada, Janssen, Lundbeck, Lundbeck Institute, Medscape, Michael Smith Foundation for Health Research, MITACS, Neurotorium, Ontario Brain Institute, Otsuka, Pfizer and Viatris, Shanghai Mental Health Center, Sunnybrook Health Sciences Center, Unity Health, Vancouver Coastal Health Research Institute, and the Vancouver General Hospital-University of British Columbia Hospital Foundation. TC has received grant funding from the Michael Smith Foundation for Health Research, and the National Research Council Canada.

Multimedia Appendix 1
Interview scripts.
[PDF File (Adobe PDF File), 229 KB - games_v11i1e43904_app1.pdf]

Multimedia Appendix 2
Modifications made to bWell-D tasks after end-user feedback - Overview.
[PDF File (Adobe PDF File), 124 KB - games_v11i1e43904_app2.pdf]

Multimedia Appendix 3
Changes made to bWell-D - Description.
[PDF File (Adobe PDF File), 102 KB - games_v11i1e43904_app3.pdf]

Multimedia Appendix 4
Examples of modifications made to bWell-D after end-user feedback - Screenshots.
[PDF File (Adobe PDF File), 46552 KB - games_v11i1e43904_app4.pdf]

Multimedia Appendix 5
Re-interview script.
[PDF File (Adobe PDF File), 162 KB - games_v11i1e43904_app5.pdf]
References


**Abbreviations**

- **MDD:** major depressive disorder
- **VR:** virtual reality

Edited by G Eysenbach; submitted 28.10.22; peer-reviewed by V Girishan Prabhu, T Wykes, A Preston; comments to author 20.12.22; revised version received 14.01.23; accepted 21.01.23; published 07.04.23.

Please cite as:

Hernandez Hernandez ME, Michalak E, Choudhury N, Hewko M, Torres I, Menon M, Lam RW, Chakrabarty T
Co-design of a Virtual Reality Cognitive Remediation Program for Depression (bWell-D) With Patient End Users and Clinicians: Qualitative Interview Study Among Patients and Clinicians
JMIR Serious Games 2023;11:e43904
URL: https://games.jmir.org/2023/1/e43904
doi:10.2196/43904
PMID:37027

©Maria Elena Hernandez Hernandez, Erin Michalak, Nusrat Choudhury, Mark Hewko, Ivan Torres, Mahesh Menon, Raymond W Lam, Trisha Chakrabarty. Originally published in JMIR Serious Games (https://games.jmir.org), 07.04.2023. 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.
Effectiveness of a Personalized, Chess-Based Training Serious Video Game in the Treatment of Adolescents and Young Adults With Attention-Deficit/Hyperactivity Disorder: Randomized Controlled Trial

María Rodrigo-Yanguas, PhD; Marina Martín-Moratínos, MSc; Carlos González-Tardón, PhD; Fernando Sanchez-Sanchez, MSc; Ana Royuela, PhD; Marcos Bella-Fernández, MSc; Hilario Blasco-Fontecilla, MD, PhD

1Servicio de Psiquiatría, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain
2Facultad de Medicina, Universidad Autónoma de Madrid, Madrid, Spain
3Universidad de Diseño y Tecnología, Madrid, Spain
4R&D Department, Hogrefe TEA Ediciones, Madrid, Spain
5Biostatistics Unit, Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain
6Consorcio de Investigación Biomédica en Red: Epidemiología y Salud Pública, Madrid, Spain
7Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain
8Departamento de Psicología, Universidad Pontificia de Comillas, Madrid, Spain
9Consorcio de Investigación Biomédica en Red: Salud Mental, Madrid, Spain
10Ita Mental Health, Madrid, Spain

*these authors contributed equally

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Servicio de Psiquiatría, Hospital Universitario Puerta de Hierro Majadahonda
C/ Manuel de Falla s/n
Majadahonda, 28222
Spain
Phone: 34 911911690
Email: hmblasco@yahoo.es

Abstract

Background: Compared with traditional approaches, gaming strategies are promising interventions for the treatment of attention-deficit/hyperactivity disorder (ADHD). We developed a serious game, The Secret Trail of Moon (TSTM), for ADHD treatment.

Objective: The main objective of this clinical trial was to demonstrate the effectiveness of an add-on, either TSTM or Therapeutic Chess (TC), in previously optimally drug-titrated, clinically stable patients with ADHD.

Methods: This study is a prospective, unicentric, randomized clinical trial in clinically stable patients with ADHD, aged 12 to 22 years. The TSTM (n=35) and TC groups (n=34) performed 12 weekly sessions of their respective treatments. The control group (CG) patients (n=35) were called by phone every week, but they received no cognitive intervention. The primary end point was the change from baseline to end point in the parent “Behavior Rating Inventory of Executive Function-2” (BRIEF-2; patients’ parents) in the per-protocol population (31 serious videogame: 24 TC and 34 CG).

Results: Our study failed to probe clear-cut improvements in the global score of the BRIEF-2. However, the TC group showed improvements in measures of emotional control, emotional regulation, and inattention. The TSTM group showed improvements in measures of emotional regulation, inattention, and school context.

Conclusions: TSTM and TC did not improve executive function symptoms, but they improved ADHD symptomatology related to emotional regulation. Further studies with bigger samples are required to confirm these preliminary findings.

Trial Registration: ClinicalTrials.gov NCT04355065; https://clinicaltrials.gov/ct2/show/NCT04355065
KEYWORDS

attention-deficit hyperactivity disorder; ADHD; serious video games; cognitive training; chess; video game; teen; young adult; game; intervention; treatment; emotional; control; regulation; attention; school; function; symptom

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is the most frequent neurobehavioral disorder, with a worldwide prevalence of 4% to 8% [1,2] and a relative risk of 3:1 against males [3]. ADHD is multicausal; about 70% is due to genetic factors, and the other 30% is due to environmental factors [4]. It is characterized by the presence of 3 core symptoms: inattention, hyperactivity, and impulsivity. These symptoms must be present in 2 or more contexts and interfere with daily functioning [5,6]. Furthermore, patients with ADHD experience a main deficit in prefrontal and anterior areas [7,8], the inferior parietal lobe, supramarginal gyrus, basal ganglia, and premotor area. Deficiencies in both executive functions [9-13] and emotional intelligence [4,14] are critically involved with ADHD.

Reference guides recommend multimodal treatment for ADHD: pharmacological intervention and behavioral therapy—behavioral training (BT) for parents and organizational training for adolescents and adults or a combination [4,15,16]. Unfortunately, multimodal treatment is not always sufficient for the treatment of ADHD symptoms [17,18]. The literature on the efficacy of cognitive behavioral therapy in ADHD is still equivocal, although some trials focused on organizational training in adolescents and adults showed promising results [19-21]. Cognitive behavioral therapy is not recommended in children younger than 10 years. Accordingly, novel strategies are warranted [22-24]. For instance, BTs are the most recommended psychological treatments for ADHD [25]. BTs are aimed at stimulating executive functions, among others [26]. Unfortunately, BTs are financially expensive and may be monotonous for adolescent patients [27]. To address these handicaps, some therapists have begun to use gaming strategies, such as (1) traditional board games like chess [28-30] and Go [31], (2) 2D video games [32-36], and (3) virtual reality serious video games [37-39]. Indeed, in a recent clinical trial including 857 children with ADHD, the patients randomized to Akili Interactive—the first videogame that can be prescribed to patients with ADHD—improved more than those who were randomized to the digital control intervention group [40]. Compared with traditional therapies, gaming strategies have several advantages, such as increased engagement, lower costs [28,41-44], and improved performance [45].

In this context, we developed The Secret Trail of Moon (TSTM), a virtual reality chess-based serious video game aimed at improving attention, impulse control, visuospatial capacity, planning, self-regulation, working memory, reasoning, and cognitive flexibility. TSTM is innovative, motivating, and customizable; it is fun, easy to understand and play, displays enjoyable graphics, and has adequate duration for most participants [39].
removed for another participant. Exclusion criteria were the following: comorbidities with autism or intellectual disability, epilepsy, or another severe clinical condition; plans to begin another clinical treatment or moving out during the following 3 months; and risk of suicide [46]. A total of 105 patients (35 per group) were recruited for the clinical trial, of which 47 (45.2%) had an inattentive ADHD subtype and 57 (54.8%) had a mixed ADHD subtype.

The participants were recruited between December 18, 2019, and November 16, 2020. The last evaluation was performed on February 6, 2021.

Randomization and Masking
To warrant clinical trial randomization, AR generated a sequence of random numbers through Epidat 4.2 software (Dirección Xeral de Saúde Pública, Xunta de Galicia). The 105 patients were randomly assigned to the 3 groups by HBF following the randomized sequence. Each randomization number was introduced inside a closed, opaque envelope. When a patient came to the hospital to participate in the trial, the inclusion and exclusion criteria were applied to them (one participant was excluded then); the patients and their parents received and signed informed consents, and after the first assessment (D0), each participant received and opened the envelope corresponding to their inclusion number, and the group assigned to that number.

Study Interventions
Visits
The number of visits depended on the group. The TSTM group made 15 in-person visits over 3 months: pre-inclusion visit, inclusion visit (D0), 12 cognitive treatment visits (1 per week) with TSTM, and one final visit (D90).

The TC and CG groups made 3 in-person visits over 3 months: pre-inclusion visit, inclusion visit (D0), and postassessment visit (D90). A total of 12 remote (web- or phone-based) visits were also made during the 3 months.

Interventions
A more detailed description of the intervention groups can be found elsewhere [46]. Basically, the interventions were different for each group, as follows:

- Cognitive intervention with TSTM (TSTM group): TSTM was constructed for 5 different cognitive training mechanisms: smasher, enigma, teka-teki, kuburi, and chess. Each session was divided into 3 blocks (2 blocks for training into 2 different game mechanics and a chess training block). Sessions were personalized in the sense that difficulty levels were adjusted, taking into account individual performance.

- Cognitive intervention with TC (TC group): in the initial visit, each patient received a flash drive containing the 12 training sessions in TC. Every session was divided into 3 parts: video tutorials, traditional chess exercises, and TC exercises.

- Control group intervention (CG group): each patient received 12 phone communications (1 per week). In these communications, they were asked about medication usage, mood, and school performance, among other issues.

Study Outcomes
Primary Outcome
The primary end point was the change from baseline to the end of treatment in the “Behavior Rating Inventory of Executive Function-2” (BRIEF-2) questionnaire filled out by the patients’ parents [49,50]. This scale consists of 63 items and 4 main indicators: emotional regulation, cognitive regulation, behavioral regulation, and executive function global index. The design of TSTM is based on cognitive training and task and aimed at improving executive functions [39]. Additionally, TC has shown a certain effect on executive functions [28].

Secondary and Safety Outcomes
Secondary measures were the ATENTO questionnaire [51], with versions for patients and for parents; the emotional intelligence inventory by the BarOn model of Emotional Quotient Inventory (BarOn EQ-i: Youth Version), with a version for younger people [52]; the ADHD rating scale V (ADHD-V); the quantitative scale by Swanson, Nolan, and Pelham (SNAP-IV) [53]; the Conners’ Parent Rating Scale (CPRS-HI) [54]; and the Conners’ Continuous Performance Test 3 (CPT-3) for computers [55]. In addition, the TSTM group filled out the Udvål für Kliniske Undersølgester (UKU) questionnaire [56,57].

Statistics
Sample-Size Calculation
Sample sizes were calculated for 80% of statistical power, assuming 2-tailed Student t tests for comparison between groups, with a 1:1:1 proportion, assuming a 20% loss. In each group, 35 patients were included (a total of 105 patients).

Analyses
Analyses were performed for two types of populations: analyses for treatment intention (ATI) and analysis per protocol (PP). In ATI data were analyzed according to the assigned treatment for each patient. In PP, data were analyzed according to the treatment that each patient completed.

A descriptive analysis was performed for qualitative variables using absolute and relative frequencies. For quantitative variables, the median and the quartiles 1 and 3 were used as descriptive statistics.

To compare the groups TSTM versus CG and TC versus CG, 2 series of analyses were performed in parallel. To compare the basal and final differences, the Kruskal-Wallis test was performed, as the normality assumption for ANOVAs and t tests was generally violated. The signification level was fixed at P=.10 mainly due to the pandemic situation and the early stage of this trial [58]. Subsequently, for those variables reaching statistical significance in Kruskal-Wallis tests, pairwise Mann-Whitney tests were performed to compare pairs of experimental groups. The effect sizes were estimated through $\eta^2$ for Kruskal-Wallis tests and $\rho^2$ for Mann-Whitney tests [59]. The statistical package SPSS 26 (IBM Corp) was used to perform the tests. Ultimately, the study was powered to detect an effect of Cohen $d \geq 0.52$ in two-way comparisons.
Results

Patient Inclusion and Randomization

A total of 105 patients agreed to participate in the trial and signed a reported consent form. Ultimately, 1 patient was excluded for having a previously unreported history of epilepsy. The other 104 patients were randomized to one of 3 groups: TSTM (n=35), TC (n=34), and CG (n=35). Figure 1 shows a flow diagram describing in detail the number of participants in each trial stage.

![Flow diagram](image.png)

Figure 1. CONSORT Flow Diagram for the number of participants in each stage of the trial.

Baseline Characteristics

In the PP population at D90, participants were male in 71.9% (64/89) of the cases (TSTM group: 27/31, 87.1%; TC group: 19/24, 79.2%; and CG: 18/34, 53.1%). The mean ages for the groups were 14.7 (SD 2.4) for TSTM, 13.8 (SD 1.9) for TC, and 14.8 (SD 2.4) for CG. Table 1 and Table 2 show all the sociodemographic and baseline variables.
Table 1. Sociodemographic and baseline variables for analysis for treatment intention (ATI). Italicized P values are significant.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>ATI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSTM(^a) (n=35)</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
<td>8 (22.86)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>14.7 (2.4)</td>
</tr>
<tr>
<td>Birthplace (Spain), n (%)</td>
<td>34 (97.14)</td>
</tr>
<tr>
<td>Adoption, n (%)</td>
<td></td>
</tr>
<tr>
<td>Adopted</td>
<td>2 (5.71)</td>
</tr>
<tr>
<td>Shelter home</td>
<td>0 (0)</td>
</tr>
<tr>
<td>ADHD(^e) diagnosed, n (%)</td>
<td></td>
</tr>
<tr>
<td>Inattentive</td>
<td>19 (54.29)</td>
</tr>
<tr>
<td>Combined</td>
<td>16 (45.71)</td>
</tr>
<tr>
<td>Medication, n (%)</td>
<td></td>
</tr>
<tr>
<td>Elvanse</td>
<td>24 (68.57)</td>
</tr>
<tr>
<td>Medikinet</td>
<td>7 (20)</td>
</tr>
<tr>
<td>Equasym</td>
<td>3 (8.57)</td>
</tr>
<tr>
<td>Rubricon</td>
<td>1 (2.86)</td>
</tr>
<tr>
<td>Stratera</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Intuniv</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Concerta</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^a\)TSTM: The Secret Trail of Moon.
\(^b\)TC: Therapeutic Chess.
\(^c\)CG: control group.
\(^d\)N/A: not applicable.
\(^e\)ADHD: attention-deficit/hyperactivity disorder.
The BRIEF-2 questionnaire shows that the TSTM group performed better than the TC and CG groups in “working memory,” “planning and organization,” and “cognitive regulation” scales, the groups being approximately equal in the rest of the BRIEF-2 scales (Tables S1 and S2 in Multimedia Appendix 1 present descriptive baseline characteristics).

For both primary and secondary outcomes, we calculated individual pre-post differences and used these differences as the dependent variables in our tests.

**Primary Outcomes**

The goal of this study was ultimately to compare the changes between D0 and D90 in every group as well as the differences among groups. The main measurement for testing the differences were the BRIEF-2 scores, according to the indications already presented elsewhere [46]. The groups were compared through Kruskal-Wallis tests.

ATI showed improvements for the TC group compared to the CG group in “emotional control” and “emotional regulation” (P < .10) scores. PP also showed a significant difference in “emotional control” between the TC and CG groups (P < .10). None of the rest of the comparisons reached a statistical significance. All tests and significance levels are shown in Tables 3 and Table 4.

### Table 2. Sociodemographic and baseline variables for analysis per protocol (PP). Italicized P values are significant.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSTM&lt;sup&gt;a&lt;/sup&gt; (n=31)</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
<td>4 (12.9)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>14.3 (2.246)</td>
</tr>
<tr>
<td>Birthplace (Spain), n (%)</td>
<td>30 (96.8)</td>
</tr>
<tr>
<td>Adoption, n (%)</td>
<td></td>
</tr>
<tr>
<td>Adopted</td>
<td>1 (3.2)</td>
</tr>
<tr>
<td>Shelter home</td>
<td>0 (0)</td>
</tr>
<tr>
<td>ADHD&lt;sup&gt;e&lt;/sup&gt; diagnosed, n (%)</td>
<td></td>
</tr>
<tr>
<td>Inattentive</td>
<td>17 (54.8)</td>
</tr>
<tr>
<td>Combined</td>
<td>14 (45.2)</td>
</tr>
<tr>
<td>Medication, n (%)</td>
<td></td>
</tr>
<tr>
<td>Elvanse</td>
<td>21 (67.74)</td>
</tr>
<tr>
<td>Medikinet</td>
<td>6 (19.35)</td>
</tr>
<tr>
<td>Equasym</td>
<td>3 (9.68)</td>
</tr>
<tr>
<td>Rubricon</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Stratera</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Concerta</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>TSTM: The Secret Trail of Moon.

<sup>b</sup>TC: Therapeutic Chess.

<sup>c</sup>CG: control group.

<sup>d</sup>N/A: not applicable.

<sup>e</sup>ADHD: attention-deficit/hyperactivity disorder.
Table 3. Kruskal-Wallis tests for primary outcomes (Behavior Rating Inventory of Executive Function-2 [BRIEF-2] scales).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>ATI(^a) (n=100)</th>
<th>PP(^b) (n=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kruskal-Wallis</td>
<td>(\varepsilon^2)</td>
</tr>
<tr>
<td>Inhibition</td>
<td>0.693</td>
<td>0.007</td>
</tr>
<tr>
<td>Self-supervision</td>
<td>0.255</td>
<td>0.003</td>
</tr>
<tr>
<td>Flexibility</td>
<td>1.425</td>
<td>0.014</td>
</tr>
<tr>
<td>Emotional control</td>
<td>4.388</td>
<td>0.044</td>
</tr>
<tr>
<td>Initiative</td>
<td>0.114</td>
<td>0.001</td>
</tr>
<tr>
<td>Working memory</td>
<td>0.610</td>
<td>0.006</td>
</tr>
<tr>
<td>Planning and organization</td>
<td>0.617</td>
<td>0.006</td>
</tr>
<tr>
<td>Task supervision</td>
<td>3.055</td>
<td>0.031</td>
</tr>
<tr>
<td>Material organization</td>
<td>0.301</td>
<td>0.003</td>
</tr>
</tbody>
</table>

\(^a\) ATI: analyses for treatment intention.  
\(^b\) PP: per protocol.

Table 4. Kruskal-Wallis tests for BRIEF-2 Indexes.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>ATI(^a) (n=100)</th>
<th>PP(^b) (n=89)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kruskal-Wallis</td>
<td>(\varepsilon^2)</td>
</tr>
<tr>
<td>Behavioral regulation</td>
<td>0.496</td>
<td>0.005</td>
</tr>
<tr>
<td>Emotional regulation</td>
<td>3.090</td>
<td>0.031</td>
</tr>
<tr>
<td>Cognitive regulation</td>
<td>1.808</td>
<td>0.018</td>
</tr>
<tr>
<td>Global executive function</td>
<td>1.435</td>
<td>0.014</td>
</tr>
</tbody>
</table>

\(^a\) ATI: analyses for treatment intention.  
\(^b\) PP: per protocol.

Secondary Outcomes

Analyses for secondary outcomes were performed in the PP population. Table 5 shows all the tests and significance levels for the secondary outcomes. Table 5 highlights those variables that reached significance in Kruskal-Wallis tests and the pairwise comparisons between the three trial groups.
Table 5. Kruskal-Wallis tests for secondary outcomes (only for PP). The italicized P values are significant.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Kruskal-Wallis</th>
<th>$\epsilon^2$</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAP-IV(^a): inattention</td>
<td>2.859</td>
<td>0.032</td>
<td>.24</td>
</tr>
<tr>
<td>SNAP-IV: hyperactivity</td>
<td>0.231</td>
<td>0.003</td>
<td>.89</td>
</tr>
<tr>
<td>SNAP-IV: total</td>
<td>1.218</td>
<td>0.014</td>
<td>.54</td>
</tr>
<tr>
<td>ATENTO (self-report): inattention</td>
<td>2.262</td>
<td>0.026</td>
<td>.32</td>
</tr>
<tr>
<td>ATENTO (self-report): hyperactivity</td>
<td>1.300</td>
<td>0.015</td>
<td>.52</td>
</tr>
<tr>
<td>ATENTO (self-report): attentional control</td>
<td>1.785</td>
<td>0.020</td>
<td>.41</td>
</tr>
<tr>
<td>ATENTO (self-report): behavioral regulation</td>
<td>0.975</td>
<td>0.011</td>
<td>.61</td>
</tr>
<tr>
<td>ATENTO (self-report): emotional regulation</td>
<td>3.507</td>
<td>0.040</td>
<td>.17</td>
</tr>
<tr>
<td>ATENTO (self-report): working memory</td>
<td>0.439</td>
<td>0.005</td>
<td>.80</td>
</tr>
<tr>
<td>ATENTO (self-report): flexibility</td>
<td>2.137</td>
<td>0.024</td>
<td>.34</td>
</tr>
<tr>
<td>ATENTO (self-report): planning and organization</td>
<td>2.776</td>
<td>0.032</td>
<td>.25</td>
</tr>
<tr>
<td>ATENTO (self-report): time orientation</td>
<td>7.170</td>
<td>0.081</td>
<td>.03</td>
</tr>
<tr>
<td>ATENTO (self-report): behavioral problems</td>
<td>1.942</td>
<td>0.022</td>
<td>.38</td>
</tr>
<tr>
<td>ATENTO (self-report): sleeping problems</td>
<td>0.741</td>
<td>0.008</td>
<td>.69</td>
</tr>
<tr>
<td>ATENTO (self-report): family context</td>
<td>0.750</td>
<td>0.009</td>
<td>.69</td>
</tr>
<tr>
<td>ATENTO (self-report): school context</td>
<td>5.283</td>
<td>0.060</td>
<td>.07</td>
</tr>
<tr>
<td>ATENTO (self-report): social context</td>
<td>0.431</td>
<td>0.005</td>
<td>.81</td>
</tr>
<tr>
<td>ATENTO (family): inattention</td>
<td>0.219</td>
<td>0.002</td>
<td>.90</td>
</tr>
<tr>
<td>ATENTO (family): hyperactivity</td>
<td>0.272</td>
<td>0.003</td>
<td>.87</td>
</tr>
<tr>
<td>ATENTO (family): attentional control</td>
<td>1.236</td>
<td>0.014</td>
<td>.54</td>
</tr>
<tr>
<td>ATENTO (family): behavioral regulation</td>
<td>0.380</td>
<td>0.004</td>
<td>.83</td>
</tr>
<tr>
<td>ATENTO (family): emotional regulation</td>
<td>0.799</td>
<td>0.009</td>
<td>.67</td>
</tr>
<tr>
<td>ATENTO (family): working memory</td>
<td>1.337</td>
<td>0.015</td>
<td>.51</td>
</tr>
<tr>
<td>ATENTO (family): flexibility</td>
<td>2.902</td>
<td>0.033</td>
<td>.23</td>
</tr>
<tr>
<td>ATENTO (family): planning and organization</td>
<td>2.563</td>
<td>0.029</td>
<td>.28</td>
</tr>
<tr>
<td>ATENTO (family): time orientation</td>
<td>3.521</td>
<td>0.040</td>
<td>.17</td>
</tr>
<tr>
<td>ATENTO (family): behavioral problems</td>
<td>3.097</td>
<td>0.035</td>
<td>.21</td>
</tr>
<tr>
<td>ATENTO (family): sleeping problems</td>
<td>1.146</td>
<td>0.013</td>
<td>.56</td>
</tr>
<tr>
<td>ATENTO (family): family context</td>
<td>2.640</td>
<td>0.030</td>
<td>.27</td>
</tr>
<tr>
<td>ATENTO (family): social context</td>
<td>2.230</td>
<td>0.025</td>
<td>.33</td>
</tr>
<tr>
<td>EQ-i:YV(^b): positive impression</td>
<td>5.826</td>
<td>0.066</td>
<td>.05</td>
</tr>
<tr>
<td>EQ-i:YV: mood</td>
<td>2.307</td>
<td>0.026</td>
<td>.32</td>
</tr>
<tr>
<td>EQ-i:YV: total emotional intelligence</td>
<td>3.827</td>
<td>0.043</td>
<td>.15</td>
</tr>
<tr>
<td>EQ-i:YV: intrapersonal</td>
<td>0.696</td>
<td>0.008</td>
<td>.71</td>
</tr>
<tr>
<td>EQ-i:YV: interpersonal</td>
<td>5.002</td>
<td>0.057</td>
<td>.08</td>
</tr>
<tr>
<td>EQ-i:YV: adaptability</td>
<td>1.909</td>
<td>0.022</td>
<td>.39</td>
</tr>
<tr>
<td>EQ-i:YV: stress management</td>
<td>2.772</td>
<td>0.032</td>
<td>.25</td>
</tr>
<tr>
<td>CPT-3(^c) response style</td>
<td>2.905</td>
<td>0.033</td>
<td>.23</td>
</tr>
<tr>
<td>CPT-3 detectability</td>
<td>5.483</td>
<td>0.062</td>
<td>.06</td>
</tr>
<tr>
<td>CPT-3 omissions</td>
<td>2.552</td>
<td>0.029</td>
<td>.28</td>
</tr>
<tr>
<td>CPT-3 commissions</td>
<td>2.369</td>
<td>0.027</td>
<td>.31</td>
</tr>
</tbody>
</table>
According to the protocol study of our add-on clinical trial [46], scores on ADHD-5, CPRS, SNAP-IV, ATENTO, BarOn EQ-i:YV, and CPT-3 tests are considered secondary end point analyses. The groups were compared by means of Kruskal-Wallis tests and, subsequently, Mann-Whitney tests. Neither SNAP-IV nor CPRS-HI scores showed significant differences between CG and any of the other groups. ATENTO (parents’ reports) showed no significant differences between the CG and the other groups.

### Table 6. Pairwise Mann-Whitney tests for those variables that reached significance in Tables 3 to 5. The italicized *P* values are significant.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mann-Whitney Test</th>
<th>η²</th>
<th>P value</th>
<th>Direction of difference</th>
<th>Mann-Whitney Test</th>
<th>η²</th>
<th>P value</th>
<th>Direction of difference</th>
<th>Mann-Whitney Test</th>
<th>η²</th>
<th>P value</th>
<th>Direction of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATENTO (self-report): time orientation</td>
<td>219</td>
<td>0.068</td>
<td>.01</td>
<td>TC&gt;TCSTM</td>
<td>519</td>
<td>&lt;0.001</td>
<td>.94</td>
<td>__d</td>
<td>273</td>
<td>0.058</td>
<td>.02</td>
<td>TC&gt;CG</td>
</tr>
<tr>
<td>ATENTO (self-report): school context</td>
<td>315</td>
<td>0.007</td>
<td>.43</td>
<td>—</td>
<td>357.5</td>
<td>0.055</td>
<td>.03</td>
<td>TSTM&gt;CG</td>
<td>328.5</td>
<td>0.023</td>
<td>.16</td>
<td>—</td>
</tr>
<tr>
<td>EQ-i:YV: positive impression</td>
<td>261</td>
<td>0.034</td>
<td>.08</td>
<td>TC&gt;TSTM</td>
<td>351</td>
<td>0.060</td>
<td>.02</td>
<td>CG&gt;TSTM</td>
<td>402</td>
<td>0.001</td>
<td>.78</td>
<td>—</td>
</tr>
<tr>
<td>Conners’ CPT-3f detectability</td>
<td>219</td>
<td>0.068</td>
<td>.01</td>
<td>TC&gt;TSTM</td>
<td>428.5</td>
<td>0.018</td>
<td>.20</td>
<td>—</td>
<td>356.5</td>
<td>0.011</td>
<td>.33</td>
<td>—</td>
</tr>
<tr>
<td>Conners’ CPT-3f perseverations</td>
<td>234</td>
<td>0.056</td>
<td>.03</td>
<td>TC&gt;TSTM</td>
<td>427.5</td>
<td>0.019</td>
<td>.19</td>
<td>—</td>
<td>348.5</td>
<td>0.015</td>
<td>.25</td>
<td>—</td>
</tr>
<tr>
<td>Conners’ CPT-3f reaction time variability (hits)</td>
<td>224</td>
<td>0.063</td>
<td>.02</td>
<td>TC&gt;TSTM</td>
<td>402.5</td>
<td>0.029</td>
<td>.11</td>
<td>—</td>
<td>335</td>
<td>0.019</td>
<td>.19</td>
<td>—</td>
</tr>
<tr>
<td>Conners’ CPT-3f block change</td>
<td>290</td>
<td>0.017</td>
<td>.22</td>
<td>—</td>
<td>370.5</td>
<td>0.047</td>
<td>.04</td>
<td>CG&gt;TSTM</td>
<td>265.5</td>
<td>0.064</td>
<td>.02</td>
<td>CG&gt;TC</td>
</tr>
<tr>
<td>Conners’ CPT-3f inter stimulus change</td>
<td>215.5</td>
<td>0.071</td>
<td>.01</td>
<td>TC&gt;TSTM</td>
<td>334</td>
<td>0.071</td>
<td>.01</td>
<td>CG&gt;TSTM</td>
<td>359</td>
<td>0.010</td>
<td>.35</td>
<td>—</td>
</tr>
</tbody>
</table>

### Notes:

- **a**SNAP: Swanson, Nolan, and Pelham.
- **b**EQ-i:YV Emotional Quotient Inventory: Youth Version.
- **c**CPT-3: Continuous Performance Test 3.
- **d**CPRS: Conners’ Parents Rating Scale.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Kruskal-Wallis</th>
<th>η²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT-3 perseverations</td>
<td>5.207</td>
<td>0.059</td>
<td>.07</td>
</tr>
<tr>
<td>CPT-3 reaction times (hits)</td>
<td>4.139</td>
<td>0.047</td>
<td>.13</td>
</tr>
<tr>
<td>CPT-3 reaction time variability (hits)</td>
<td>6.465</td>
<td>0.073</td>
<td>.04</td>
</tr>
<tr>
<td>CPT-3 variability</td>
<td>3.340</td>
<td>0.038</td>
<td>.19</td>
</tr>
<tr>
<td>CPT-3 block change</td>
<td>7.605</td>
<td>0.086</td>
<td>.02</td>
</tr>
<tr>
<td>CPT-3 inter stimulus change</td>
<td>9.007</td>
<td>0.102</td>
<td>.01</td>
</tr>
<tr>
<td>CPRSd</td>
<td>0.884</td>
<td>0.010</td>
<td>.64</td>
</tr>
</tbody>
</table>
ATENTO (self-report) scores showed a significant difference between the TSTM and CG groups in “school context” (P<.10). Additionally, the TC group showed a statistically significant difference with the CG in “time orientation” measures. None of the other comparisons reached statistical significance.

BarOn EQ-i:YV scores showed statistically significant differences between the TSTM and CG groups in “positive impression” on the interpersonal scales (P<.10). None of the remaining comparisons reached statistical significance.

CPT-3 scores showed differential improvements in “block change” and “inter stimulus change” in the TSTM group compared with the control group and differences in improvements in “block change” between the TC and control groups (P<.10). None of the other comparisons reached statistical significance.

**Discussion**

**Principal Findings**

This prospective, unicentric, randomized study in clinically drug-stable children with ADHD did not confirm our hypothesis that using an augmentation strategy (either TSTM or TC) may improve executive functions as measured by the BRIEF-2 total scores. An exception was the improvement in emotional control and regulation as measured by the BRIEF-2 in the TC group. Furthermore, significant pre-post symptom reductions were found for several scales measuring secondary end points, and they differ depending on the type of intervention (as explained in the following sections).

**TSTM**

We found no improvements in executive functioning after using TSTM. This finding is consistent with literature using other serious video games such as ACTIVATE [60] or Plan-It Commander [61]. However, this last video game demonstrated improvements in working memory and planification skills, which have not been observed in our study. In addition, a recent review found that some video games may be beneficial for executive function training in patients with ADHD [62].

However, we found that the TSTM group showed better posttreatment scores in core ADHD symptoms, in particular inattentive symptoms, compared to the CG, mainly in self-reported tests. Moreover, parents reported improvements in social and school skills, consistent with results coming from studies using other serious video games like Plan-It Commander [61] and ThinkRx [63].

But the most relevant finding is that the TSTM group, compared to the CG, showed some improvement in some emotional domains. Both self-reports and parent reports provided better scores on scales related to emotional intelligence, emotional regulation, and performance in school context compared to the CG, which is in keeping with some other literature [64-67]. Better than a purely cognitive tool, the TSTM video game may serve as a way to train and improve self-regulation and social skills through facing challenging problems and competitions with a social component.

**TC**

As was the case with TSTM, we did not find any improvement in the TC group when compared with the CG, with the remarkable exception of emotional control and regulation. This negative finding is in contradiction with several studies reporting improvements in executive functioning using a therapy based on board games (eg, chess and Go) [31]. Demily et al [68] reported that patients with schizophrenia who played chess for just 10 sessions improved executive functioning (z=-3.41). Finally, a recent systematic review and meta-analysis suggested that some of these board games may improve self-efficacy, attitudes, social functioning, and executive functions [69]. Our negative findings might be explained either by differences in the studied populations (the studies were devoted to patients with diagnoses apart from ADHD) or that TC was implemented electronically.

However, core ADHD symptoms benefited from TC. In particular, inattentive symptoms improved most in the TC group when compared with the CG. This result is consistent with that obtained in a pilot study using chess in children with ADHD [28]. Furthermore, as was the case with TSTM, the greatest improvements obtained by TC training were in the emotional and interpersonal spheres. Significantly better improvements were shown in emotional intelligence, emotional regulation and control, as well as interpersonal scales, as measured by the BRIEF-2 and BarOn EQ-i:YV scores. Our findings are in keeping with previous studies reporting that chess training has shown good results in emotional intelligence development [30,70].

**Strengths and Limitations**

The main strength of this study is the use of 2 gaming strategies for treating ADHD, following an add-on strategy with clinically stable patients; one of them was disruptive—the use of a serious video game (TSTM)—and the other (TC) was more traditional, thus allowing for testing the effectiveness of each tool in executive functions and ADHD core symptoms. In contrast, the main limitation was the irruption of the COVID-19 pandemic, which impacted the clinical trial, as some patients randomized to the TSTM group had more difficulty completing the cognitive training sessions. Another limitation was the use of just one TSTM session per week in the TSTM group, which may have limited the effectiveness of this cognitive training. Accordingly, we plan to include a 2-3 session-per-week strategy in the next clinical trial using TSTM, scheduled to begin in April 2023. Furthermore, the evaluation was not blind and was based on clinical judgment and either self-report or parental reports. However, previous work states that teachers, rather than parents, tend to be the most reliable external observers of ADHD symptoms in children [71]. Furthermore, we used the CPT-3 to have an objective measure of our results. Another limitation was the difference between the ATI and the PP populations. Samples sizes for the groups were not very different, but a potential survival bias may have conditioned our results. It is possible that adherence to treatment was conditioned by the differential attractiveness of the video game and TC, that is, children who like video games or chess were more likely to complete the trial, and that could bias the results for the PP
population. Indeed, in the PP analyses, fewer patients finished the TC intervention in the TC group. It is possible that only those patients with ADHD who were motivated or who liked chess finished the treatment, thus biasing our results in the TC arm. Another difference is the alternate location of the TSTM treatment, administered in our hospital, compared to the TC treatment, self-administered at the patients’ homes. Moreover, a possible limitation is the use of drug-stable samples; perhaps the effects of playing serious video games or TC may be more prominent in patients without a stable pharmacological treatment. Furthermore, to have clinical samples as representative and general as possible, we recruited participants in a wide age range, but our samples were not large enough to allow for analyses across developmental stages; future research should study in depth the differential effects across ages. Finally, the lack of a long-term follow-up approach kept us from studying the maintenance of any beneficial effect over time.

Conclusions
Neither TSTM nor TC produced a significant improvement in the executive functioning of patients with ADHD who were clinically stable on ADHD drugs. However, both TSTM and TC produced a significant improvement in some ADHD symptoms, particularly in emotional regulation areas. Further studies are warranted to extend these preliminary but motivating findings.

Acknowledgments
We thankfully acknowledge Lorraine Maw, MA, for her support in proofreading the manuscript.

This research was partially funded by the Spanish National Project (grant RTI2018-101857-B-I00), an iPFIS (Contratos Predoctorales de Formación e Investigación en Salud) research contract (Instituto de Salud Carlos III IFI16/00039), a grant from the Fundación para la Innovación y la Prospectiva en Salud en España, and an intensification of research activity for health professionals grant (2019) from the Puerta de Hierro Health Research Institute-Segovia de Arana.

Authors’ Contributions
MRY and MMM conceptualized the study and were in charge of the software, investigation, data curation, writing the original draft, and funding acquisition. MMM also reviewed and edited the manuscript. CGT contributed to the conceptualization and the software.

FSS provided resources. AR conducted the methodology, formal analysis, and data curation. MBF contributed to the formal analysis as well as writing, reviewing, and editing the manuscript. HBF conceptualized the study, provided resources; contributed to the writing, reviewing, and editing the manuscript; supervised the study; and was also in charge of funding acquisition.

Conflicts of Interest
In the last 24 months, HBF has received lecture fees from Shire. He is the principal investigator (PI) of an iPFIS (Contratos Predoctorales de Formación e Investigación en Salud) research contract (Instituto de Salud Carlos III; IFI16/00039) and the co-PI of a MINECO research grant (RTI2018-101857-B-I00); he is the recipient of (1) a grant from the Fundación para la Innovación y la Prospectiva en Salud en España and (2) an intensification grant from Puerta de Hierro Health Research Institute-Segovia de Arana; involved in 2 clinical trials (Mensia Koala, NEWROFEED Study; ESKETSUI2002); and the cofounder and CEO of Haglaia Solutions. He is also an employee and member of the advisory board of Ita Salud Mental (Korian). MMM is the recipient of a grant from the Centro para el Desarrollo Tecnológico y la Innovación, funded by the European Fund for Regional Development (IDI-20180701; file 00107278). The remaining authors do not have any conflicts of interest.

Editorial notice: This randomized study was only retrospectively registered, as the authors thought it to be registered, but after starting the trial realized that it had not been registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. 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
Descriptive baseline characteristics.
[DOCX File, 23 KB - games_v11i1e39874_app1.docx ]

Multimedia Appendix 2
CONSORT e-Health checklist.
[PDF File (Adobe PDF File), 75 KB - games_v11i1e39874_app2.pdf ]

References


Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ATI: analyses for treatment intention
BRIEF-2: Behavior Rating Inventory of Executive Function-2
CG: control group
CPRS: Conners’ Parent Rating Scale
CPT-3: Continuous Performance Test 3
EQ-i: YV Emotional Quotient Inventory: Youth Version
PP: per protocol
TC: Therapeutic Chess
TSTM: The Secret Trail of Moon
UKU: Udvalg für Kliniske Undersøgester
Original Paper

Serious Game for the Screening of Central Auditory Processing Disorder in School-Age Children: Development and Validation Study

Ana-Martá Gabaldón-Pérez, MSc; María Dolón-Poza, MSc; Martina Eckert, PhD; Nuria Máximo-Bocanegra, PhD; María-Luisa Martín-Ruiz, PhD; Iván Pau De La Cruz, PhD

1Grupo de Investigación Innovación Tecnológica para las Personas (InnoTep), Departamento de Ingeniería Telemática y Electrónica, ETSIS de Telecomunicación, Campus Sur, Universidad Politécnica de Madrid, Madrid, Spain
2Centro de Investigación en Tecnologías Software y Sistemas Multimedia para la Sostenibilidad (CITSEM), Universidad Politécnica de Madrid, Madrid, Spain
3Department of Physiotherapy, Occupational Therapy, Rehabilitation, and Physical Medicine, Faculty of Health Sciences, Universidad Rey Juan Carlos, Madrid, Spain

Corresponding Author:
María-Luisa Martín-Ruiz, PhD
Grupo de Investigación Innovación Tecnológica para las Personas (InnoTep)
Departamento de Ingeniería Telemática y Electrónica
ETSIS de Telecomunicación, Campus Sur, Universidad Politécnica de Madrid
Calle Nikola Tesla
Madrid, 28031
Spain
Phone: 34 910673316
Email: marialuisa.martinr@upm.es

Abstract

Background: Currently, many central auditory processing disorder screening tests are available for children, and serious games (SGs) are frequently used as a tool for the diagnosis of different neural deficits and disorders in health care. However, it has not been possible to find a proposal that unifies both ideas. In addition, the validation and improvement of SGs, in general, does not take into account the player-game interaction, thus omitting valuable information about the playability and usability of the game.

Objective: This study presented Amalia’s Planet, a game conceived for use in school environments, which allows a first assessment of a child through their performance of the proposed tasks related to different aspects of auditory performance. In addition, the game defines a series of events in relation to the execution of the tasks, which were evaluated for the subsequent optimization of its performance and the improvement of its usability.

Methods: Using screening tools based on the use of SG technologies, a total of 87 school-age children were evaluated to test the various hypotheses proposed in this study. By grouping users according to whether they had personal history of hearing pathologies, the discriminant power, playability, and usability of the final solution were examined using traditional statistical techniques and process mining (PM) algorithms.

Results: With a confidence level of 80% for test 2 ($P=.19$), there was no statistical evidence to reject the null hypothesis that a player’s performance is affected by whether the player had a previous auditory pathology. Furthermore, the tool allowed the screening of 2 players initially categorized as healthy because of their low level of performance in the tests and the similarity of their behavior with that of the group of children with a previous pathology. With regard to the validation of the proposed solution, the use of PM techniques made it possible to detect the existence of events that lasted too long, which can lead to player frustration, and to discover small structural flaws in the game.

Conclusions: SGs seem to be an appropriate tool for the screening of children at risk of central auditory processing disorder. Moreover, the set of PM techniques provides a reliable source of information regarding the playability and usability of the solution to the development team, allowing its continuous optimization.

(JMIR Serious Games 2023;11:e40284) doi:10.2196/40284
KEYWORDS
serious games; central auditory processing disorder; process mining; screening; school environment; children

Introduction

Goal of This Study

Owing to their high degree of neuroplasticity, children and young people can be highly influenced by the stimuli they receive from the environment around them. Given the nature of central auditory processing disorder (CAPD), it is believed that the inclusion of auditory stimulation may lead to functional changes in the individual’s central auditory nervous system [1]. More specifically, “children come to experience greater and more complex demands on central auditory processing as they face more academically, intellectually and linguistically demanding challenges” [2].

Before the therapeutic treatment of CAPD, it is necessary to carry out a diagnostic or screening process so that the presence of the pathology can be detected, thus facilitating a prognosis and the suppression or reduction of the limitations that it may cause in the patient. Among the diagnostic and rehabilitation tools for different pathologies are the well-known serious games (SGs). According to Piaget [3], “games form a part of the child’s intelligence, because they represent the functional or reproductive assimilation of the reality.” However, despite the existence of numerous screening tests for CAPD [4] and the more than palpable efficacy of SGs in multiple settings [5], it has not been possible to find a project in which these 2 techniques are unified to support the screening of individuals who may have the disorder in question.

In the field of SGs, there is a wide variety of methodologies that guide the development and validation of such tools according to their purpose [6-8]. Despite this, most game evaluation and optimization processes focus on game performance, omitting a complete analysis of the player-game interaction, which includes other relevant aspects such as playability, usability, and game understanding. At this point, it is important to highlight the fact that the term usability refers to “the extent to which a system (product or service) can be used to achieve the goals with effectiveness, efficiency, and satisfaction in a specified use context” [9].

Process mining (PM) techniques can be used to deepen our knowledge about this player-game interaction. PM allows for the discovery of patterns and tendencies, even without previous models, enabling and improving the comprehension of real processes by recording events. “Process Mining aims to discover, monitor, and improve real processes by extracting knowledge from event logs readily available in today’s information systems” [10].

In view of this, the main objective of this research was to develop and validate a screening tool for CAPD based on the SGs technology, supporting the diagnosis of the disorder in school-age children. Amalia’s Planet was conceived as a game for use in school environments, which allows a first assessment of a child through their achievement of the different proposed tasks related to auditory discrimination, lateralization and localization of sounds, and different aspects of auditory performance. In addition, the game defines a series of events in relation to the execution of the tasks, which were evaluated using PM techniques for the subsequent optimization of its performance and improvement of its usability [11].

It is important to note that throughout this paper, SG is defined according to the study by Tarja et al [12], which stated that an SG is “any game whose main purpose goes beyond the mere entertainment.”

Background

The American Speech-Language-Hearing Association uses the term CAPD to refer to deficits in the neural processing of auditory information in the central auditory nervous system not due to higher-order language or cognition, as demonstrated by poor performance in one or more of the skills used to preserve, refine, analyze, modify, organize, and interpret information from the auditory periphery [13].

As indicated in the technical reports published by the Working Group on Auditory Processing Disorders, the process of training the abilities affected by CAPD should be implemented as early as possible, taking advantage of the plasticity and cortical reorganization characteristics of the human brain [13,14]. This principle of action could have a much more positive impact on children because the maturation of the cerebral cortex and creation of neural interconnections occur during childhood. Therefore, from the perspective of the consequent therapeutic process, diagnostic testing or screening for CAPD in school-age children is highly desirable. Undiagnosed children often present with learning disabilities, occasionally mistaken for other pathologies simply because auditory processing problems were not considered.

Currently, there are different tools that allow the observation and subsequent screening and diagnosis of CAPD in individuals. Focusing on children, the Hearing in Noise Test for Children [15] lists a total of 130 sentences to determine the quality of information received by an individual in a noise context. Similarly, the Pediatric Speech Intelligibility test proposes a context of overlapping sounds, with relevant and distracting information being uttered in the same or opposite ears [16,17]. The Test Everyday Attention for Children [18] focuses on the assessment of different aspects of an individual that determine their attention levels [19]. Despite their proven efficacy, all these tests tend to be time-consuming and depend on a high degree of specialization on the part of professionals, which restricts their accessibility for the general child population.

Through the use of SGs, a greater number of children can simultaneously take different screening tests, requiring only the company of their relatives or educators and being able to carry out the activity in a familiar environment. Added to these facilities is the fact that because evaluation is performed using a game, the child is stimulated during the process. Finally, another notable advantage of the use of these technologies is that they allow for their evaluation and consequent evolution,
making it possible to add different criteria or informative factors in the future as well as to optimize the existing ones. Despite these advantages and the existence of some games focused on the therapeutic process of CAPD, it has not been possible to find a game with these characteristics for the screening of children with CAPD [20-23].

The self-evaluation feature and potential evolution of the games are made possible by the validation process carried out after the use of the games by the end users. During this process, the characteristics of the tool, such as its playability, usability, and discriminating power, must be evaluated. Owing to the lack of publications related to the generation of CAPD-screening SGs for children, different methodologies proposed for the development and validation of this type of technology in the health field have been reviewed. However, most of them base the validation of the tool on performance indicators without assessing aspects of the individual or the tool that may alter the value of these parameters. In the first place, Verschueren et al [24] generated a framework for the conception of SGs for health. This framework was created to ensure that the final product is relevant, theoretically driven, and evidence based through 5 stages. More centered on auditory impairments, Cano et al [25] carried out an exhaustive study of different methodologies for the conception of SGs. According to the authors, all methodologies lacked a register of the player’s learning in a quantitative manner. After learning about the successes and failures of these methodologies, they created the metodología para la concepción de juegos serios para niños con discapacidad auditiva (MECONESIS) methodology. MECONESIS is a methodology consisting of 4 stages, namely analysis, preproduction, production, postproduction, and is based on 6 models, namely the analysis, user, pedagogical objective, task, scenario, and validation models.

Among the methodologies that take into account the player-game interaction, the one proposed by Serrano et al [26] in 2014 stands out. The authors proposed a “framework to improve evaluation in educational games.” On the basis of learning analytics and game analytics techniques, the authors suggest a Learning Analytic Model and Learning Analytic System, in which the authors point to some events carried out by players on a graphical interface, the evolution of game stages, and even some logic events. Despite all these sources of information, no tool or model capable of analyzing the collected information was provided.

PM, as a set of techniques designed to extract information from existing data in activity log files, draws on the so-called event logs or game events from different systems, allows them to be cleaned, establishes contextual relationships, and presents them graphically to facilitate decision-making [27]. This tool provides great advantages to those who use it, among which 3 are worth highlighting in this study [9,28]:

1. PM is based on real events; therefore, the information collected and analyzed is a reflection of reality. It is obvious that, in general, when extracting information for reaching a series of conclusions, the aim is to gather information that is as representative of reality as possible. However, traditional statistical techniques, for example, try to extract such information from a sample of a specific population (sampling techniques). That is, a sampling behavior is generalized to the entire population. In addition, statistics have techniques for the elimination of so-called missing values and outliers so that their presence does not affect the final conclusions that can be reached. For its part, PM works with all the information collected in the event logs, thanks to its high degree of automation. Thus, behaviors are neither generalized nor modified. Atypical behaviors are taken into account and compared with standard or normal behaviors.

2. It includes a set of defined algorithms, which allow complex events to be analyzed and conclusions to be drawn to improve the system. That is, PM provides the capability to analyze and optimize the system. Every system, tool, and environment in its first phase of development and throughout its use must be evaluated and updated for its correct functioning. The aim is not only to optimize the system in terms of performance but also to adapt the system to the end user so that it provides the desired usability.

3. It presents the analyzed information in a visual manner, making it easily understandable by any user, to facilitate the decision-making process.

All these advantages make PM the ideal tool for the future optimization and improvement of the SG proposed in this study.

Methods

Overview

Amalia’s Planet was conceived as an SG for use in the school environment, which allows a first detection of children at risk of CAPD through their completion of the different tasks proposed. Therefore, through the game and the proposed activities, schools will be able to perform screening in cases in which there is no suspicion of CAPD or in which such suspicion is not entirely clear. In addition, the game defines a series of events in relation to the execution of the tasks, which were evaluated using PM techniques for the subsequent optimization of its performance and improvement of its usability.

From the aforementioned statement, 2 conclusions regarding the approach to the solution are evident. First, a multidisciplinary team must be assembled to cover the wide variety of fields of expertise required for the correct deployment of the solution. In this regard, the team was composed of 3 engineers; 1 PM expert; 1 occupational therapist; 2 audiologists; and 1 ear, nose, and throat specialist. Second, the development of the solution must be approached from 2 points of view: in terms of the health field, there is a need to develop a screening tool with discriminant power for children at risk of CAPD and, in terms of new technologies and data analysis, there is a need to develop of a game that provides relevant information related to different aspects of playability and usability to allow for its evaluation and subsequent optimization.

Nonetheless, although the existence of the aforementioned 2 approaches (health and technological or analytical) is clear, it should not be forgotten that the proposed solution is one and has to cover all the requirements that have been imposed in each
of the 2 approaches. This is why, later in this paper, both approaches are mentioned simultaneously.

Ethics Approval
This study was approved by the Ethics Committee of the Rey Juan Carlos University (reference 1203201805218). The trial was conducted in accordance with the Declaration of Helsinki, as amended in October 2013 by the 64th World Medical Association General Assembly [29].

Amalia’s Planet for Screening Children at Risk of CAPD
As there is no unified screening method for CAPD, current standard screening methods for CAPD include observing the listening process, recording the patient’s behavior, and performing a series of tests that examine the patient’s auditory function performance [30]. In general, individuals with CAPD have difficulty discriminating sounds, impaired ability to determine the stimulated ear and location of the sound source, and problems processing nonverbal acoustic signals and recognizing the order or pattern of presentation of certain stimuli [31]. These and other conditions experienced by affected individuals should serve as indicators for the tool so that it can detect the presence of the pathology.

In addition, the proposed solution is oriented toward its use among school-age children. As children are the end users for whom the solution is developed, it should take into account not only the characteristics of the disorder to be studied but also the characteristics of the age range of the users. In other words, the game should present an attractive and a motivating environment for children, making them increase their commitment when interacting with the tool. Figure 1 presents an outline of the process of developing the game as a screening tool.

Figure 1. Phase 1: deployment of a screening tool for school-aged children from a health care perspective. PM: process mining; SG: serious game.

Amalia’s Planet is set, as the name suggests, on a planet consisting of multiple modules (“habitats” or “islands”) of different types. Each “habitat” is home to a different type of animal and incorporates multiple mini-games, all created with a simple cartoon design, making them easily accessible to primary-school children. The animals chosen are those that normally form herds or flocks, and they emit different sounds that are collectively handled as “noise.” This makes it possible to establish noisy backgrounds that produce sound and visual distractions for the child and thus forces them to train listening skills where they have difficulties.

Located in Antarctica, the habitat presented in this paper focuses on the assessment of certain skills affected by various CAPD-related pathologies. More specifically, the skills assessed in the first module are auditory discrimination, sound localization, auditory memory, and auditory separation in noise. For this, the habitat has a series of mini-games, which sequentially assess the mentioned skills. Each of these mini-games provides a series of stimuli and rewards throughout the game so that the player’s motivation is maintained.

As the first mini-game in the Antarctic habitat, the child has to find a stray baby penguin of a herd in a wide area guided only by the sound generated by the animal (auditory localization). The original positions assigned to both the penguin and player are randomly chosen by the game each time the game starts (Figure 2). As a measure of the individual’s performance in this test, the time taken to find the little penguin is collected. Once the child finds the little penguin, this achievement is celebrated by the herd through an applause given to the player.
Once the player passes the first test, the second mini-game starts automatically (Figure 3). This is an auditory memory game. Playing with the head penguin, the child must be able to memorize the sequences of sounds produced by the penguin. Each of the sequences emitted by the head penguin can comprise a series of sounds that the player would have to remember: clapping the wings, twinkling an eye, rattling the beak, and stomping a foot. For the player to become familiar with the aforementioned sounds, the penguin first exposes them together with an animated movement (visual aid). Once the child is deemed to have the necessary information to play the game independently, the penguin hides within its huddle (disappearance of the visual aid) and proceeds to emit sequences of sounds in a given order. The player must remember both the sounds heard and the order in which they were produced to reproduce them once the penguin finishes emitting the sounds.

With regard to the length of the sequences, they become more complex as the child passes the tests. The number of sounds to be distinguished and the length of the sequence emitted by the animal gradually increase in 3 different levels of difficulty, each of which consists of 3 sublevels (Figure 4). In addition, each sublevel provides the individual with 3 opportunities to pass, and if all 3 opportunities are exhausted without success, the sublevel and its corresponding level shall be considered as not passed. If the player manages to pass any of the levels, applause starts to ring out, and the penguin congratulates the player so that they feel motivated and satisfied with the task performed. If the player does not manage to complete any of the tasks correctly, the penguin encourages them to continue playing (thereby trying to avoid frustration and abandonment) and repeats the sequence of sounds that the child has to recreate.

In addition, the game allows the configuration of this auditory memory test for the evaluation of auditory discrimination skills and sound separation in noise. If desired, the penguin colony hiding the main penguin will emit noise to distract the player, demanding higher auditory abilities. This background sound is adjustable for each player and can vary between 0 dB and 30 dB.
**Amalia’s Planet as Its Self-evaluator**

The postproduction stage “corresponds to the evaluation of the serious game, where an evaluation model is proposed, taking into account 2 roles of evaluators, the end user and the expert” [25]. It is at this stage that the previously mentioned PM techniques were introduced as a novel tool for the validation of the proposed solution. However, as reflected in Figure 1, it was at the preproduction stage that the event logs that would be analyzed later have to be determined.

Because information was collected in the system logs, it allowed the definition of events by the team, which provided flexibility when addressing problems. Looking at the structure of the game (Figure 4) and with the idea of knowing the behavior of the players throughout the game, the team defined the events...
according to the result that the player had obtained in each sublevel. At this point, 3 possible outcomes were defined:

1. Passing the sublevel: the player reproduces all the sounds of the sequence in the correct order.
2. Omitting the sublevel: the player does not reproduce all the sounds in the sequence. This result was considered a failure.
3. Failing the sublevel: the player either does not reproduce all the sounds of the sequence correctly or does not reproduce the sounds in the correct order.

Thus, the traces of the event logs consisted of events that report the path followed by each player throughout their progress in the game. This definition of events allowed us to know the basic behaviors of the players, such as the number of attempts a player needed to pass a level and the nature of the level (omission or failure). In addition, it allowed us to recognize patterns of interest that could occur, such as more frequent cycles, traces, or paths among players that are not easily perceived using conventional statistical tools. Figure 5 schematically shows the tool validation process proposed by the development team.

![validation.png](attachment:validation.png)

**Figure 5.** Phase 2: Proposed Amalia’s Planet validation process. SG: serious game.

The information provided by PM about the system’s functioning, performance, and use as a tool for achieving specific objectives can serve as a source of input for various purposes, which highlights PM’s capacity in this regard. Not only can conclusions be drawn about the achievement of the objectives by users, but hypotheses that could not be validated or were poorly formulated at the outset can also be modified because of the new information obtained. In fact, new hypotheses can be put forward on the basis of behaviors collected in the event logs, which had previously not been taken into account because of a lack of knowledge.

As shown in Figure 5, the validation process of the hypotheses raised affects previous steps, such as the definition of the game events or the very conception and design of the game. This iterative and dynamic process is what allows the solution to be fine-tuned. The ultimate goal of this loop process is that, based on the information self-generated by the game, the system itself is nourished, updated, and adapted to the demands and characteristics of the player, as well as to the process of analysis of this information presented to the team of experts.

**Experiment and Hypothesis**

Amalia’s Planet was tested twice in a school in the Community of Madrid, Spain.

Information was collected through interviews with parents about their children’s medical history related to previous hearing pathologies, such as previous hearing infections (otitis, drains, and secretions) and ear surgeries, and whether their children had a diagnosis of attention-deficit/hyperactivity disorder (ADHD) or learning difficulties [32,33]. Hereafter, this group of conditions will be referred to as previous hearing pathologies.

To evaluate the achievement of the discriminating objective of the game (screening tool), the interdisciplinary team proposed the following hypotheses:

1. There is a difference in gameplay performance between children with previous hearing pathologies and children without previous hearing pathologies.
2. “Healthy” players learn faster. Children who have not had any of the aforementioned conditions are expected to improve their gameplay performance faster than children with a diagnosed pathology.

3. Children learn while they play. Regardless of whether children have had a hearing condition, it is interesting to know whether the game improves their trained skills.

To test the formulated hypotheses, a statistical study of the difference in means between the 2 groups of children (with and without previous hearing difficulties) was conducted. In addition, different scatter plots were studied with respect to measures of central tendency with the aim of finding atypical behaviors that reveal the presence of individuals at risk. These hypotheses are assessed later in this paper.

Results

Sample Summary

The sample size was determined in such a way that, according to the central limit theorem, the statistical distribution of each of the populations considered for the study (children with previous hearing pathologies, children without previous hearing pathologies, distribution of children by sex, etc) was correctly reflected in the collected data [34]. Thus, theoretically, a sample size of 50 children was determined for each sample group (n>30), with a total sample of 100 children in the study. Nevertheless, owing to the limited availability of users of interest for the study and some problems in data collection using the tool (3 players were excluded), the final size of the sample used for this study was 87 children.

The sample consisted of children of both sexes (50 female and 37 male children) aged between 5 and 14 (mean 8.87, SD 1.74; median 9) years. Most of the players (76/87, 87%) were aged between 7 and 11 years, of whom 55% (42/76) were female and 45% (34/76) were male, maintaining approximately the initial distribution of the sample.

Regarding the pathology line, without taking into account sex or age, 45% (39/87) of the players had or used to have at least 1 pathology among those contemplated in the study, resulting in 55% (48/87) of “healthy” participants. The sex ratio was as follows: 57% (50/87) female and 43% (37/87) male. Health information obtained through the parent questionnaire revealed the following classification by previously identified pathology: (1) 55% (48/87) of the sample had a history of oitis, of whom 33% (16/48) had major complications, although only 28% (24/87) had undergone a surgery for drainage; (2) 3% (3/87) were diagnosed with ADHD; and (3) a similar percentage 3% (3/87) had learning disorders.

The sample included children with a history of hearing pathologies and children without a history of hearing pathologies. Hence, focusing on the players with some previous pathology (39/87, 45%), there were 49% (19/39) of players with just 1 pathology, 31% (12/39) of players with 2 different pathologies, and 21% (8/39) of players with up to 3 types of pathologies. Overall, 92% (36/39) of the players with at least 1 pathology had or used to have oitis.

Among the 19 children with just 1 affliction, 84% (n=16) had oitis (a percentage close to the prevalence of oitis in the total population of 87 children, which is estimated at 80% [35]), and 11% (n=2) had ADHD, making it the second most frequent pathology appearing individually. Generally, suppurations and learning disorders appeared to be accompanied by at least 1 other type of pathology in all cases. The most common combination of afflictions was made up of oitis and suppurations, which was present in 11 children (55% of the players with at least 2 pathologies). In terms of sex, the percentage of female players with some kind of affliction was close to the percentage of male players with some kind of affliction: of the 50 female players, 22 (44%) had a pathology, and of the 37 male players, 17 (46%) had a pathology.

Overview of the Obtained Health Results

Results Obtained in Test 1

For this first test, a sample of 59 children was available. Among the 59 children, 29 (49%) had previous hearing pathologies, with a total of 14 (48%) male and 15 (52%) female children. On the other hand, a total of 51% (30/59) children were part of the group of children who were not diagnosed with hearing pathologies, with a total of 50% (15/30) male and 50% (15/30) female children.

The percentage of players with afflictions who managed to finish the game successfully was lower than the percentage of “healthy” players (11/29, 38% of players with some affliction against 16/30, 53% of “healthy” players). To determine whether this difference in the level of playing between the different groups was statistically significant, a chi-square test was performed, obtaining a P value of .24. Therefore, whether the child had or used to have a hearing pathology did not seem to affect their ability to successfully complete the game. However, it was interesting for the team to answer the following question: is it possible that some of the children initially classified as “healthy” players may in fact have one of the pathologies considered in the study or a different one and have not been diagnosed? Taking the following variables as references, throughout the collection of the information, it was observed that there were certain players initially cataloged as “healthy” whose level of gaming stood out compared with the level of playing of their peers:

- TOTAL_ATTEMPTS: the number of attempts the player has made throughout their game
- TOTAL_ACCURACIES: the number of successes the player has produced throughout the game
- FAILURE_TOTAL: the total number of failures the player has faced during the game
- TOTAL_SUBS: the number of sounds omitted by the player throughout the game

Figures 6A to 6D show (from left to right) the existence of players initially classified as “healthy” who, compared with their group, showed atypical behaviors in their results, particularly, the players who did not manage to pass the game, presenting a large number of errors or omissions throughout their game (ID 30 and ID 74).
Results Obtained in Test 2

For the second test, a sample of 70 children was available. Among the 70 children, 28 (40%) had previous hearing pathologies, with a total of 13 (46%) male and 15 (54%) female children. On the other hand, a total of 60% (42/70) children were part of the group of children who were not diagnosed with hearing pathologies, with a total of 45% (19/42) male and 55% (21/42) female children.

In this second test, most players managed to pass all the levels of the game. There was a clear difference between the percentage of children with previous pathologies who managed to finish the game completely and the percentage of children without previous pathologies who managed to finish the game completely (26/42, 62% of the players without pathology against 12/28, 40% of the players with some type of diagnosed pathology). To determine whether this difference was statistically significant, a chi-square test was performed, obtaining a $P$ value of .19. Therefore, at a confidence level of 80%, there was no statistical evidence to reject the hypothesis that having the afflictions in question affects the ability to complete the game successfully.

To determine whether the players who stood out in test 1 maintained the trend in the second test, the variables mentioned earlier were analyzed again to study their distribution in the sample. Figure 7 shows that these players maintained the same tendencies along the second trial.
Figure 7. Test 2: (A) total number of attempts for healthy players who did not complete the game, (B) total accuracies for “healthy” players who did not complete the game, (C) total number of failures for healthy players who did not complete the game, and (D) total number of omissions for healthy players who did not complete the game.

Overview of the Obtained Game Validation Results

As noted earlier, the achievement of the test objectives should not be the only measure of the test’s fitness for purpose, but aspects such as the time required to pass each of the tasks and the different event traces constituted by each player can be sources of information to be considered. To this end, the team used the ProM (Prom Software Inc) software, which is an extensible framework that supports a wide variety of PM techniques in the form of plug-ins [27]. By evaluating the different event logs defined in the previous stages, the abovementioned features can be evaluated and, if necessary, modified further by adapting the tool to its final objective.

Figure 8 depicts the duration of each of the events that were defined in the initial stage. Based on the understanding that the duration of an event is the time that elapses from the time when the “faster” player initializes it until the time when the “slower” player finishes it, this graph can reveal the existence of events that pose a greater challenge to the individual. To avoid the feeling of frustration that the player may experience, and the consequent decrease in performance, this particular feature of PM techniques allows the team to modify such game events.

Figure 8. Duration of events.

Furthermore, the study of event traces and the frequency with which they appear in the collected sample serves as a mirror of the correct functioning of the game (Figure 9). Thus, it is possible to detect structural faults in the tool, such as jumps between levels that are not consecutive, backtracking that is not allowed, and advancing to a new level without having completed the previous one. All these constituent faults can be corrected by the technical part of the development team, ensuring the correct functioning of the tool in future tests.

Finally, another graph of interest that can be obtained using the ProM software is the so-called Business Process Model Notation [27]. This graph provides information about the route that different players followed to pass the game, identifying the main events that each player went through. Furthermore, loops and cycles can be detected using this graph; thus, the team can deduce the possible learning resources of the players (Figure 10).
Discussion

Principal Findings
SGs are increasingly being used as support tools for the development and improvement of skills in individuals of different age ranges in areas such as education and health. In addition, it is nowadays demonstrated that an early diagnosis of CAPD in an individual is a desirable characteristic from the point of view of the consequent therapeutic process [36]. In other words, school-age children who are screened for CAPD may benefit greatly from the treatment of CAPD symptoms. Despite this, it is currently not possible to find a tool that brings together these 2 ideologies to minimize, or even eliminate, the afflictions that this disorder causes in those who have it.

On the basis of the conception and development of an SG, Amalia’s Planet, and its use in a real school environment, the research team was able to contrast the different hypotheses put forward in previous sections regarding the suitability of the use of SGs as screening tools for CAPD. More specifically, this study showed that once the users become familiar with the game and its rules, users who were affected by auditory pathologies related to the disorder in question perform less well on the tasks they are given. More specifically, at a confidence level of 80%, there was no statistical evidence to reject the hypothesis that
there is a difference in gameplay performance between children with hearing pathologies and children without hearing pathologies. That is, the fact that the player had any of the pathologies considered in this study directly affects their performance in the game. Moreover, the abovementioned statement only supports the hypothesis that “healthy” children learn faster than children with a diagnosis. Finally, an improvement in the performance of the players could be observed when comparing their performance in test 1 with that in test 2.

One of the main results obtained in this study is that using the proposed solution, the team was able to detect the presence of 2 users who were initially assigned to the group of “healthy” individuals owing to the lack of diagnosis but who presented a gameplay performance more similar to that of the group of at-risk individuals. Such an outcome allows family members, educators, and health professionals to pay more acute attention to these users. Such children can then be examined to determine whether their performance was altered by the presence of an auditory processing pathology or by other external factors.

At this point, it is worth highlighting the fact that, using traditional statistical techniques on extreme values, both users were eliminated from the sample, and the data were reanalyzed, the results obtained would have been even more discriminant. For both tests 1 and 2, with statistical confidence levels of 85% and 95%, respectively, there would have been no statistical evidence to reject the hypothesis about the difference in performance between “healthy” players and those who had been diagnosed.

In turn, the use of PM techniques as a tool for evaluating the suitability of the solution to the problem posed allowed the team to develop different methods for improving the game. In particular, by studying the duration of the game events, the team detected some phases of the game that can generate frustration in the players, thus reducing their level of attention and, consequently, their performance. In addition, different learning resources used by the players were detected. By studying cycles and loops along the player’s traces, learning resources based on failure and omission that were used by the player to subsequently complete the levels could be detected. Finally, the team was able to detect jumps between nonconsecutive levels of the game. All these features to be improved in the tool could be addressed in future versions, thanks to the validation of the proposed solution through PM techniques.

**Limitations of the Study and Future Works**

After testing the developed tool, the team was able to detect a series of aspects to improve in the application. First, at the time of data analysis, a lack of information was detected from a total of 3 users. This lack of information due to an error in the deployment of the solution led to the exclusion of these players from the final analysis. Moreover, in the first game in which the child has to find the baby penguin, a bug was detected. Here, some players had to cross a bridge to reach their goal. However, some of them fell from the bridge into the water and required the help of adults to continue, adding time to their markers and lowering their score. Moreover, thanks to the use of PM tools, events in which the dynamic followed in the game was not established could be detected a priori. Thus, in Figure 10, a jump from the initial event to the final event without going through any intermediate states can be detected. In addition, in other cases, player games were detected in which the game “jumps” from one level to another, ignoring intermediate sublevels.

Finally, events that were “excessively” long compared with others were detected. These events will have to be reviewed, as they can suppose a defective performance on the part of the tool or a high level of complexity for the players involved.

In addition to the revision and correction of the mentioned errors, the team of authors is currently developing new “habitats” for the Amalia’s Planet ecosystem, which covers new aspects related to CAPD screening and were not included in this first version.

**Conclusions**

CAPD is a disorder that is difficult to detect during childhood, and its early detection is a challenge to be solved in the near future using quick and easy-to-apply screening tools. To our knowledge, this is the first study that unifies the ideology of a screening test for CAPD in school-age children and the use of SGs to support the diagnosis and consequent treatment of CAPD. Despite the need for further testing of the usability of the tool and the improvement of certain functional aspects of the tool, the development team was able to obtain results with statistical evidence that support the use of Amalia’s Planet as a screening tool for children at risk for CAPD. Amalia’s Planet can be expanded to include games that assess other forms of afflictions experienced by children with the disorder under study.

**Acknowledgments**

The authors would like to thank the team of nursery and primary teachers at the Colegio de Amanecer (Madrid, Spain) for their collaboration in performing the experiment presented here.

**Conflicts of Interest**

None declared.

**References**

1. Bellis TJ. Auditory processing disorders: it's not just kids who have them. Hear J 2003 May;56(5):10-19 [FREE Full text] [doi: 10.1097/01.HJ.0000293883.42025.ca]


Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CAPD: central auditory processing disorder
MECONESIS: metodología para la concepción de juegos serios para niños con discapacidad auditiva
PM: process mining
SG: serious game

©Ana-Marta Gabaldón-Pérez, María Dolón-Poza, Martina Eckert, Nuria Máximo-Bocanegra, Maria-Luisa Martín-Ruiz, Iván Pau De La Cruz. Originally published in JMIR Serious Games (https://games.jmir.org), 26.04.2023. 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.
Original Paper

Serious Games Based on Cognitive Bias Modification and Learned Helplessness Paradigms for the Treatment of Depression: Design and Acceptability Study

Arka Ghosh¹, PhD; Jagriti Agnihotri¹, BTech; Sradha Bhalotia¹, MTech; Bharat Kumar Sati¹, BTech; Latika Agarwal¹, BTech; Akash A¹, M Tech; Swastika Tandon¹, MBA; Komal Meena¹, BTech; Shreyash Raj¹, BS; Yatin Azad¹, BTech; Silky Gupta², MS; Nitin Gupta¹,²,³, PhD

¹Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, India
²Department of Cognitive Science, Indian Institute of Technology Kanpur, Kanpur, India
³Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, Kanpur, India

Corresponding Author:
Nitin Gupta, PhD
Department of Biological Sciences and Bioengineering
Indian Institute of Technology Kanpur
Kalyanpur
Kanpur, 208016
India
Phone: 91 5122594384
Email: guptan@iitk.ac.in

Abstract

Background: Depression is a debilitating mental health disorder, with a large treatment gap. Recent years have seen a surge in digital interventions to bridge this treatment gap. Most of these interventions are based on computerized cognitive behavioral therapy. Despite the efficacy of computerized cognitive behavioral therapy–based interventions, their uptake is low and dropout rates are high. Cognitive bias modification (CBM) paradigms provide a complementary approach to digital interventions for depression. However, interventions based on CBM paradigms have been reported to be repetitive and boring.

Objective: In this paper, we described the conceptualization, design, and acceptability of serious games based on CBM paradigms and the learned helplessness paradigm.

Methods: We searched the literature for CBM paradigms that were shown to be effective in reducing depressive symptoms. For each of the CBM paradigms, we ideated how to create a game so that the gameplay was engaging while the active therapeutic component remained unchanged.

Results: We developed 5 serious games based on the CBM paradigms and the learned helplessness paradigm. The games include various core elements of gamification, such as goals, challenges, feedback, rewards, progress, and fun. Overall, the games received positive acceptability ratings from 15 users.

Conclusions: These games may help improve the effectiveness and engagement levels of computerized interventions for depression.

(JMIR Serious Games 2023;11:e37105) doi:10.2196/37105

KEYWORDS
serious games; cognitive bias modification; learned helplessness; depression; digital intervention; mobile phone

Introduction

Background
Depressive disorders are a leading cause of disability worldwide, affecting >264 million people [1]. Antidepressant medications and cognitive behavioral therapy (CBT) are currently the most effective treatment modalities for depression [2-4], but there is a large treatment gap owing to the lack of accessibility, high cost of therapy, and social stigma associated with the visits to a psychiatric clinic [5-7]. In the past 2 decades, digital interventions have been gaining traction. The most widely
studied digital intervention for depression is computerized CBT (cCBT). Although cCBT is an effective form of treatment [8-10], a recent meta-analysis has demonstrated that guided cCBT is less acceptable than being on a waiting list [11]. There is a need for other digital treatment modalities.

The cognitive theory of depression posits that information processing—attention, interpretation, and memory—is negatively biased in depression [12,13]. This biased information processing has been implicated in the onset and maintenance of depression [14]. Experimental studies have corroborated the cognitive model by providing empirical support for self-referential information processing, attention bias, interpretation bias, and memory bias [15-19]. Psychological training paradigms, known as cognitive bias modification (CBM) paradigms, have been used to modify these cognitive biases in subthreshold, clinical, and remitted populations [20-28], albeit with mixed results. Despite the mixed results and small effect sizes [29,30], this field promises novel treatment approaches and warrants further research.

The previously mentioned evaluations of CBM paradigms have been mostly performed in laboratory settings; these paradigms have received much less attention than cCBT as a potential treatment approach for depression outside the laboratory [31-33]. CBM paradigms tend to have low user engagement owing to the repetitiveness of the paradigms [34-37] and a lack of credibility in some cases [34,38]. In a digital intervention, in which the user needs to engage with a paradigm without any external supervision, the paradigm’s ability to engage becomes the determining factor for the success of the intervention.

One way to make the CBM paradigms more engaging to users is by developing serious games based on the paradigms [39]. A game serves as an interactive medium that provides game designers with considerable control in creating engaging experiences and can reduce the attrition rate in digital interventions [40]. Serious games have been evaluated for multiple conditions, including posttraumatic stress disorder [41,42], autism spectrum disorder [43-45], attention-deficit/hyperactivity disorder [46], cognitive functioning [47,48], alcohol use disorder [49,50], trait anxiety [51,52], etc. The results of a meta-analysis that evaluated the effectiveness of serious games demonstrated a moderate effect size of serious games for reducing psychiatric disorder–related symptoms compared with no intervention controls [53]. In a review of 4 studies that evaluated CBM interventions based on serious games, only 2 studies reported their gamified interventions to be effective [54].

Serious games have been developed for various aspects of depression. Bespoke serious games have been created to diagnose depression [55,56], treat depression [57-63], and describe the experience of depression either metaphorically [64] or literally [65]. Off-the-shelf commercial games have also been used to target depression [66]. There are multiple serious games based on different therapeutic techniques—CBT [67-69], solution-focused therapy [70], and interpersonal therapy [68]—that do not target depression specifically but can be useful. There is only 1 serious game based on a CBM paradigm targeting depression, but it was found to be ineffective [62].

Serious game–based interventions have been shown to be acceptable to the users overall [57,60,67-69]. A recent review, however, suggests that the effectiveness and acceptability data are not convincing enough yet to warrant clinical adoption [71].

Objective
In this study, we selected specific CBM paradigms that had received empirical support for targeting cognitive biases underlying depression and appeared suitable for conversion to games. Then, we described the design and development of these games in detail. Furthermore, we developed 1 serious game based on the learned helplessness theory.

Methods
Embedding Training Paradigms Into Serious Games
We searched the literature for known CBM paradigms and selected those that have shown promise in reducing depressive symptoms and were amenable to conversion to games. Our team discussed multiple game ideas for each paradigm. While thinking of the ideas, the primary aim was to make the gameplay engaging while keeping the active therapeutic component unchanged. After multiple sessions on ideations, the team selected the most promising game idea for each paradigm. Once the team agreed on a game idea, minimally viable prototypes were created and tested within the team. These minimally viable prototypes were improved upon to create the final versions by using the feedback obtained from informal tests performed in a larger group of approximately 10 friends and colleagues.

Software Development
We used Angular 8 (Google LLC) and JavaScript (ECMAScript 2018) for creating the front end of the games and Django-rest-framework 3.9.2 (Encode OSS Ltd) and Django 2.2 (Django Software Foundation) for creating the backend of the games. Two games required the detection of the swipe gesture on the screen, for which we used a readily available library, hammer.js. For 1 game that required the simulation of gravity and collision in the game environment, we used the Phaser.js framework. The games were optimized for smartphone screens and were also compatible with tablets, laptops, and desktop devices. We created the graphical elements required for the game using Adobe Illustrator (Adobe Inc). Musical elements required for some of the games were collected from open web resources. The negative attention bias training game required the combination of 20 rhymes (Rock a bye baby, Wheels on the bus, etc) and famous compositions (Ode to joy, Für Elise, etc). The musical notations for these rhymes and compositions were collected from various YouTube channels. A few of them were readily available in the scientific pitch notation. For the others, we deduced the musical notations from the simple piano versions of these rhymes and compositions available on YouTube.

Feedback on the Games
To obtain a preliminary sense of the acceptability of the games, we asked 15 pilot users (n=8, 53% male and n=7, 47% female) to provide feedback on the 3 aspects of the games. The users were recruited from the Indian Institute of Technology, Kanpur.
community, by word of mouth. Potential users who expressed interest were sent an email detailing the procedure to participate in the study and to provide feedback. They were also sent an informed consent form via email. They were required to include their name and signature in the informed consent form and email it back to the study team. These users rated the games on three aspects: (1) The instructions on how to play the game were clear, (2) The game was fun to play, and (3) The purpose of the game was clear. The users rated these aspects on a 5-point scale: (1) strongly agree, (2) agree, (3) neutral, (4) disagree, and (5) strongly disagree. These ratings were mapped to scores of 2, 1, 0, −1, and −2, respectively. For the game for learned helplessness, we asked the users an additional question, In this game, you were presented with some unsolvable puzzles and some solvable puzzles. After you completed the game, the logic of the game was presented to you. After playing the game and reading the explanations, which of the following statements are true (tick all that apply)?

The following options were provided: (1) I have an understanding of learned helplessness; (2) I felt angry or frustrated while I was stuck with the puzzles; (3) After the logic of the game was explained, my frustration or anger reduced; and (4) After the logic of the game was explained, my frustration or anger did not reduce. The feedback was collected as a preliminary pilot for a larger study.

Ethics Approval
Ethical clearance was provided by the Institutional Ethics Committee of the Indian Institute of Technology Kanpur (IITK/IEC/2019-20/II/4).

Analysis
We calculated the average value of the feedback received on the different aspects of the game. We also checked whether the feedback received was significantly different from 0 using the Wilcoxon signed rank test. Python (version 3.10.5; Python Software Foundation) and MATLAB R2020b (Mathworks) were used for data analysis.

Results
Design Philosophy
Our main design objectives were to make the games effective by using evidence-based paradigms and engaging by incorporating 6 out of the 7 core elements of gamification. The evidence-based paradigm used for the individual games is described in individual sections throughout the paper. CBM paradigms have been reported to be boring and repetitive [34-37]. We hypothesized that making the games fun and engaging would increase the likelihood of spontaneous gameplay and, consequently, increase the CBM dosage without external supervision. Studies have reported that the rationale behind some CBM interventions is not apparent to the users, which can contribute to the lack of engagement [34,38]. To tackle this, we added a brief section called Science behind the game to each game.

Individuals experiencing an episode of depression have decreased cognitive abilities [72]. Considering this, we designed the gameplay to be easy and intuitive, such that anyone with some experience in using smartphone apps and a basic familiarity with the English language could play the games. We also kept the hardware requirements for the games minimal so that they could be played on low-end smartphones. At the beginning of the games, we gave simple in-game instructions or in-game tutorials. To avoid an artificial limit on the practice of CBM, we designed the games to be never-ending (except for the game for learned helplessness, in which the design required a definite ending). The difficulty level of each game adapted to the skill level of the user to create the flow experience [73]. Apart from making the games intrinsically engaging, we also added some extrinsic motivators to play the games, such as badges and scores, or the sensory experience of a familiar musical tune. The badges served as short-term rewards for playing the games, and the scores served as in-game currency, which could be exchanged for lives or hints in the games. We describe how our games are based on the elements of gamification in the later section titled, How game-like are our games?

Game for Automatic Interpretation Bias Modification
People with depression have an automatic negative interpretation bias [74]. Cowden Hindash et al [21,75,76] used a modified word sentence association paradigm for participants with dysphoria to confirm and modify the automatic negative interpretation bias. The bias modification paradigm also showed a near-transfer effect on the Scrambled Sentence Task [77] and a far-transfer effect of increased resiliency on a laboratory stressor [21].

In the task of modified word sentence association paradigm for participants with dysphoria, a self-relevant ambiguous sentence (My mom called me to tell me the news) is shown for a short duration, followed by a positive or negative word. The user is asked whether the word is related to the sentence. The user can answer Yes (related) or No (unrelated). The user is trained, via text-based feedback, to associate positive words with ambiguous sentences; the feedback is Correct if the answer is Yes for a positive word or No for a negative word; otherwise, the feedback is Incorrect. We converted this paradigm into a serious game in the following manner: The game displays an alphabet grid. The user’s goal is to find the words (that can be formed by connecting letters in the grid; Figure 1A) that form the ambiguous sentence. The game screen includes several empty bars, equal to the number of words to be found. As the user finds a word, a bar gets filled to provide the users with a sense of progress in the game and indicate the number of words found. Once the user finds 50% of the words, we reveal the full ambiguous sentence. The user can answer by clicking on Yes (related) or No (unrelated) buttons. Quicker answers are given higher rewards to encourage the users to answer with the first thought they have. An incorrect answer receives 0 points. To prevent the users from clicking Yes for all positive words and No for all negative words, we used decoy sentence-word pairs, in which the sentence was unambiguous, and the negative word was related, while the positive word was unrelated [78].
The users earn reward points by finding words in the grid. Valid English words that are not parts of the sentence also earn points. The points, apart from being a short-term reward, also serve as an in-game currency to buy extra time or hints. Apart from points, the users also receive other short-term rewards in the form of bronze, silver, and gold badges for completing 4, 10, and 26 levels, respectively.

The game adapts to the user’s level of competence by varying the number of words hidden and allotted time per difficulty level (Figure S1 in Multimedia Appendix 1). To help the user in learning to play the game, the initial 4 levels of the game keep the words of the sentence fully visible in a jumbled order and give the user 150 seconds (in the first 2 levels) or 120 seconds (in the next 2 levels) to find the words in the grid. A bar, ranging from Easy to Hard, indicates the game’s current difficulty level.

There are 42 levels (sentence-word pairs) in the game. The difficulty of each level depends on the number of words to be found and the complexity of the grid. These, in turn, depend on the length of the sentence and the length of the longest word. The levels are arranged to gradually increase the difficulty level. After the completion of all the levels, the game loops back to the first level.

**Game for Executive Control Training**

Rumination, a core factor in depression, affects the duration [79] and intensity [80] of depressive episodes. Multiple studies have established that depressive rumination is associated with an inability to inhibit the processing of emotional information [81,82]. Executive control is one mechanism that regulates the ability to inhibit emotional information in healthy individuals [83-85]. Cohen et al [86] hypothesized that the difficulty in using executive control might be the factor behind impaired inhibition in ruminators. Cohen et al [86] designed a paradigm to train individuals to exert executive control before exposure to an emotional stimulus, followed by a discrimination task. They observed that individuals trained with this paradigm were less likely to engage in a state of rumination [86] and more likely to use reappraisal—an effective emotion regulation technique—more frequently and efficiently [87]. On the basis of this executive control training paradigm [87], we designed a 2D platform game to resemble the classic video game, Super Mario Bros. (Figure 2). The goal of the user is to keep a running avatar alive by jumping over obstacles (sitting or jumping frogs; Figure 2A) or pits (Figure 2B). The pits are of different types: some requiring a single jump, some requiring a double jump, and some including floating platforms in between (Figure 2C). The floating platforms are also of 2 kinds—some are stationary when the avatar lands on them, whereas others start to fall upon the landing of the avatar. The avatar sometimes passes through an underground tunnel, and on returning to the ground level, the background landscape changes to maintain a diversity of visuals. An in-game tutorial is shown to make the user familiar with the game’s rules and controls. On a keyboard-based device, the arrow keys can be used to play the game. On a touchscreen device, buttons corresponding to the arrow keys are shown on the screen itself.
As time progresses, the speed of the avatar increases to increase the difficulty level. The score in the game increases over time as the avatar continues to run. Intermittently, the game screen presents coins, which can be collected by jumping. The coins serve as an in-game currency. Initially, the user is given 3 lives and 5 double jumps for free. Once the user exhausts the free double jumps, each double jump costs 10 coins. In addition, once the user exhausts the free lives, additional lives can be purchased with coins with gradually increasing difficulty: the first purchase costs 10 coins and each purchase after that costs twice the amount of the last purchase.

At predetermined points in the game, the avatar stops in front of a large pit with floating platforms. At this point, the executive control training task is shown to the user (Figure 2D). An incongruent or a congruent set of flanker arrows (Figure 2E) is shown with a 50:50 probability. The user is required to press an arrow key corresponding to the middle arrow in the set. The flanker arrows remain visible until the user responds or for a maximum of 2000 milliseconds. After the flanker set disappears, an image is shown for 100 milliseconds. The image is neutral or negative with an 80:20 probability following a congruent set and with a 20:80 probability following an incongruent set. After a further gap of 50 milliseconds, during which no stimulus is shown, a green or red circle is shown with a 50:50 probability (Figure 2F). The user is required to respond according to the color of the circle (up arrow for green and down arrow for red). The circle remains visible until the user responds or for a maximum of 2000 milliseconds (Figure S2 in Multimedia Appendix 1). These 3 steps—including the arrows, the image, and the circle—constitute 1 round of training. The next round of similar training steps begins after a delay of 2000 milliseconds. After 3 rounds of training, the gameplay begins again and continues for approximately 30 seconds before the next batch of training begins. The instructions for choosing the correct keys are shown before each round of training. If the user presses the correct keys for both the flanker task and the colored circle task in a round, it is considered a correct response. The users are incentivized to perform the task diligently, as correct choices reward them with double jumps or additional lives. The users can also earn short-term rewards in the form of bronze, silver, and gold badges for 6, 15, and 39 correct responses, respectively.
Game for Negative Attention Bias Modification

The vulnerability model of low self-esteem and depression states that low self-esteem increases the risk of future depressive episodes [88,89]. Accordingly, improving one’s self-esteem should reduce the risk of depression. Individuals with lower self-esteem are more attentive to rejection cues, which in turn increases their tendency to interpret more and more social cues as rejecting, thus perpetuating the cycle of vigilance and low self-esteem [90,91]. Dandeneau et al [92] used an attention-training paradigm to modify this negative attention bias and observed that it led to a more resilient self-esteem against a laboratory-based rejection manipulation.

We designed a game based on the same training paradigm to overcome negative attention bias [92]. Images of human faces, one depicting a positive emotion and the others depicting negative emotions, are shown in a grid (Figure 3). The user is asked to click (or tap, when played on a touchscreen device) on the image with positive emotion as quickly as possible. Once the user clicks on the positive image, the next set of images is displayed in the same grid. To make this process fun and engaging, the game plays a musical note when the user clicks on a positive image. The musical notes, played sequentially, are taken from a popular song (or rhyme), and 1 song constitutes 1 level of the game. A click on a negative image results in an unpleasant beep. Therefore, in effect, the user can play the song by clicking on the faces with positive emotions—the more accurately they select the images with positive emotions, the fewer interruptions they hear during the song.

Figure 3. Screenshots of the game for negative attention bias training, showing (A) a $1 \times 2$ grid, (B) a $2 \times 2$ grid, and (C) a $3 \times 2$ grid of faces with different emotions.

To make the game challenging, the user is given a time limit to finish each level, calculated as $TT = N \times T$, where $N$ is the number of musical notes in the song and $T$ is the time allotted per note (initialized to 5000 milliseconds). The game adapts to the user’s performance level by dynamically varying $T$ between 1000 and 5000 milliseconds and the grid size between $1 \times 2$ (Figure 3A), $2 \times 2$ (Figure 3B), and $3 \times 2$ (Figure 3C) after the completion of each level. The algorithm of this adaptation is summarized in Figure S3 in Multimedia Appendix 1.

A bar, ranging from Easy to Hard, indicates the current difficulty level. To dissuade the user from clicking on images without paying attention, we penalize each click on a negative image by deducting 1 life (in addition to playing the unpleasant beeps). The user is given 5 free lives at the beginning of each game. Once the user runs out of lives, they have the option to replay the same level. Clicks on positive images earn points, which can be used to buy more time if the user runs out of time before the completion of a song. To give the user a sense of progress at each level, a progress bar on the screen indicates the fraction of notes in the current song that has been played.

The users also receive badges as short-term rewards. They earn bronze, silver, and gold badges by clicking on 34, 60, and 156 positive images, respectively. The game has 20 levels (songs). Once the user finishes all the songs, the game loops back to the first song.

Game for Positive Imagery Training

Negative automatic thoughts can be verbal or imaginal [93]. Holmes et al [94] developed an interpretation bias training paradigm based on positive mental imagery. Multiple studies have demonstrated that the training has a positive effect on the participants’ moods [25,94-96], although there are some exceptions [36,97]; the training also helped in targeting anhedonia [97].
We designed a serious game based on the combination of paradigms by Holmes et al [94] and Mathews and Mackintosh [98]. This game was developed as interactive fiction inspired by Zork [99]. At each level, the user is presented with an ambiguous scenario in the form of a paragraph with a blank space (Figure 4A). The user can fill in the blank with one or more words to resolve the sentence positively or negatively (e.g., You think you ______ be able to enjoy the meeting can be resolved positively by adding will or negatively by adding will not). The game evaluates the phrase written by the user by comparing it against a comprehensive list of possible positive and negative completions for each sentence (occasionally, if the phrase falls outside the list, the game responds that it could not understand the input and requests the user to write a different phrase). If the user resolves the sentence negatively, it is considered an incorrect response and the user is asked to try again (Figure 4B). If resolved positively, the game moves forward and another ambiguous scenario appears in continuation with the previous one (Figure 4C). Once the user resolves the second part of the ambiguous scenario positively, the level concludes and a new scenario appears at the next level of the game. As vivid imagination is an active element [100], users are urged to imagine the situations vividly in the first person. No hints are provided in the game; however, if the user fails to answer correctly in 3 attempts, a list of potential answers is shown from which the user can select one. To encourage vivid imagination, we did not include a time limit in this game.

**Figure 4.** Screenshots of the game for positive imagery training. (A) A scenario is shown that can be resolved positively or negatively by providing one or more words to fill in the blank. (B) If the user resolves the scenario negatively, they are asked to try again. (C) If the user resolves the scenario positively, the game moves forward and another scenario is presented.

The game is designed to be never-ending. It includes 20 levels, and after the completion of all the levels, the game loops back to the first level. The user earns 5 points for resolving a scenario positively. The user also earns a bronze, a silver, and a gold badge after positively resolving 4, 10, and 26 scenarios, respectively.

**Game for Learned Helplessness**

Learned helplessness is a laboratory model of depression that emulates multiple aspects of clinical depression in animals [101,102]. The state of learned helplessness, reached after some experiences of inescapable aversive conditions, involves a generalized self-assumption of powerlessness, thereby reducing the effort to come out of difficult situations in the future. A reformulation of the theory in terms of attribution theory has also been extended to humans: individuals with learned helplessness attribute their failures to personal, pervasive, and persistent lack of abilities and their successes to luck [103].

We designed a puzzle-based game to help the users understand the ideas of learned helplessness and attribution style. We reasoned that an explanation after a game-based transient experience of the phenomenon would be more effective than simply explaining the concept using text or videos and is more likely to help users change their self-defeating attribution bias.

In the original experiments on learned helplessness, dogs failed to escape avoidable shocks after they had unsuccessfully tried to escape unavoidable shocks [104]. Very recently, a paradigm has been proposed to test learned helplessness in humans using loud audio tones as a stressor [105]. To allow the users to appreciate this phenomenon without a physical stressor, we devised the following approach: if a user can solve a puzzle initially but gives up on the same puzzle when it is shown after
the failure to solve a hard and unrelated puzzle, the user would be able to see the parallel with learned helplessness and realize that it can happen in real life. At this point, the user is more likely to be receptive to the introspection of their own attribution bias.

We created a suite of 4 puzzle-based mini-games to achieve this objective. The first one, the circle-triangle mini-game, is a genuine puzzle game with multiple difficulty levels. The other 3 mini-games are designed to trick the user: each of them appears like a genuine puzzle game with easy puzzles in the first 1 or 2 levels but has unsolvable levels thereafter.

The user starts by playing the circle-triangle mini-game. The screen presents a pattern of squares, each containing a circle or a triangle (Figure 5A). If the user clicks on one of the squares, it flips (a circle becomes a triangle and vice versa). Simultaneously, the adjacent squares also flip (Figure 5B). The goal of the user is to change all the embedded shapes into circles. Once the user solves the puzzle, a message (Great) and auditory feedback (a pleasant ding) are provided to the user. Next, a harder level of the same mini-game is presented. This stepwise increase in the level continues until the user fails to solve the current level and clicks on a button labeled I give up (Figure 5C). When this happens, 1 of the 3 unsolvable mini-games is displayed to the user.

The 3 unsolvable mini-games are based on different types of puzzles (Figures 6A-C). For example, one is based on the popular 15 puzzle, in which a 4 × 4 grid has 1 empty cell and 15 cells containing tiles numbered 1 to 15 and the goal is to move the tiles to arrange them in increasing order. The first level shown in the mini-game is kept very simple and easily solvable (Figure 6A1), which allows the users to become familiar with the puzzle. However, unbeknownst to the user, the second and third levels are unsolvable—we set the initial configuration of tiles in these levels by slightly reordering the tiles from solvable puzzles in such a way that they could not be solved anymore but looked normal otherwise (Figure 6A2). Thus, at the second level—just after the very easy first level—the user finds it impossible to solve the puzzle and has no choice but to accept failure by clicking on the button I give up. After this, the third level of the same mini-game appears, in which the user again gives up regardless of how hard they try.

Once the user gives up in both the levels of an unsolvable mini-game, the last solved level of the circle-triangle mini-game is shown again (Figure 6 presents the flow of mini-games). If the user gives up on this level, which they have previously solved, we take the opportunity to draw a parallel with learned helplessness and the game ends with the explanation. By relating to the game, we explain the ideas of learned helplessness and attribution bias, with an emphasis on personalization and pervasiveness. We expect that once individuals can appreciate their own susceptibility, they are more likely to engage in the introspection to identify real-life situations where they fall prey to dysfunctional attribution styles. A correct understanding helps one to evaluate each situation independently and not lose motivation in all areas of life upon experiencing failure in some.

**Figure 5.** Screenshots of the circle-triangle mini-game in the game for learned helplessness. (A) The initial state in level 4. (B) An intermediate state in level 4. (C) An intermediate state in level 6.
Figure 6. The 3 unsolvable mini-games within the game for learned helplessness. (A) The goal in this mini-game is to arrange the tiles in the increasing order of the numbers. Configuration A1 is solvable, and A2 is unsolvable. (B) The goal in this mini-game is to get the butterfly to sit on all the flowers. The butterfly can be moved forward and sideways using the cursor keys on a keyboard or the swipe action on a touchscreen device. Once a butterfly sits on a flower, the flower withers away and the butterfly cannot move to a position without a flower. B1 is solvable, and B2 is unsolvable. (C) The goal in this mini-game is to insert the small red arc into the big blue arc. The user can move the solid black ball using cursor keys or swipe action to the adjacent squares (if the square contains an arc open in the direction of the ball, the ball moves inside the arc). Once the ball is inside an arc, the arc can be moved along with the ball to an adjacent square that is empty or contains a bigger arc with an opening in the direction from which the smaller arc is coming. The ball comes out of the small arc if moved in the direction in which the arc is open. C1 is solvable, and C2 is unsolvable.
If the user does not give up on the last solved level of the circle-triangle mini-game, they reach the next level within the same mini-game, and the game continues as normal. Once they give up on a higher level of the circle-triangle mini-game, they are shown a different unsolvable mini-game. This cycle of alternating between the circle-triangle mini-game and a different unsolvable mini-game continues until the user gives up on a previously solved level of the circle-triangle mini-game or completes all levels in it (the difficulty of the higher levels ensures that the latter scenario is unlikely).

The 4 puzzle-based mini-games are designed to be engaging and challenging to play. The difficulty level in the circle-triangle mini-game adapts to the competence level of the users. The screen displays the current level to provide the user with a sense of progress. The user receives a gold badge on finishing the game. Unlike other games in TreadWill, this game was designed to be played for a limited time, until the explanation of learned helplessness and the idea of the game were shown; however, the users are free to play it again if they wish.

**How Game-Like Are Our Games?**

Cugelman [106] has suggested 7 core elements of gamification. Each of the games we had designed incorporated 6 of the 7 elements, including goals, challenges, feedback, rewards, progress, and fun; the only element not included in our games is social connectivity, as the games were designed to be played individually as part of a mental health intervention. Table 1 describes how the core elements are incorporated into each game.
Table 1. Core game elements incorporated in different games.

<table>
<thead>
<tr>
<th>Game</th>
<th>Goals</th>
<th>Challenges</th>
<th>Feedback</th>
<th>Rewards</th>
<th>Progress</th>
<th>Fun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic interpretation bias modification</td>
<td>Finding words in the alphabet grid to complete the sentence</td>
<td>Finding words from the grid within the allotted time</td>
<td>Points for finding words and positive interpretation</td>
<td>Points and badges</td>
<td>• In-level progress using progress bar</td>
<td>Finding words in the alphabet grid</td>
</tr>
<tr>
<td>Executive control training</td>
<td>To keep the avatar alive as long as possible and collect coins</td>
<td>Avoid the obstacles at an increasingly faster speed</td>
<td>Points for keeping the avatar alive</td>
<td>Points and badges</td>
<td>• In-game progress via increasing avatar speed</td>
<td>Gameplay similar to Super Mario Bros.</td>
</tr>
<tr>
<td>Negative attention bias modification</td>
<td>To click on positive faces</td>
<td>To complete the song within the allotted time and avoid clicking on negative faces</td>
<td>Point and progress in level for clicking positive face and reduction in life for clicking on negative face</td>
<td>Points and badges</td>
<td>• In-level progress using progress bar</td>
<td>The music along with the game play</td>
</tr>
<tr>
<td>Positive imagery training</td>
<td>To progress in the game by providing positive resolutions</td>
<td>Finding the right word</td>
<td>Points and progress in level for positive resolutions</td>
<td>Points and badges</td>
<td>• In-level progress using progress bar</td>
<td>Imagining the self-referent situations</td>
</tr>
<tr>
<td>Learned helplessness</td>
<td>Solving the puzzles</td>
<td>Solving the puzzles</td>
<td>Positive feedback for solving a puzzle</td>
<td>Badge</td>
<td>• For the solvable game, levels are shown</td>
<td>Solving the puzzles</td>
</tr>
</tbody>
</table>

Feedback on the Games

The feedback for all games obtained from the 15 pilot users (Methods section) is summarized in Figure 8. In most cases, there was statistically significant positive feedback on the clarity of instructions, fun in gameplay, and clarity of purpose (all \( P < .05 \); Table 2), showing that the games were acceptable to the users. According to the user feedback (Figure 8A-E), the games for negative attention bias modification (Figure 8C) and positive imagery training (Figure 8D) had the clearest instructions; the game for learned helplessness (Figure 8E) was the most fun to play; and the game for positive imagery training (Figure 8D) had the clearest purpose. The game for learned helplessness was able to achieve its purpose: 11 (73%) out of 15 users said that they had a better understanding of the idea of learned helplessness after playing the game. The game was able to induce anger or frustration in 6 (40%) out of 15 participants, of which 4 (67%) of the 6 mentioned that their anger or frustration reduced once the logic of the game was explained (overall, 7/15, 47% users reported a reduction in anger or frustration). Only 2 (13%) out of 15 participants reported that their anger or frustration did not reduce once the logic of the game was explained.

Figure 8. Acceptability results of the game for (A) automatic interpretation bias modification, (B) executive control training, (C) negative attention bias modification, (D) positive imagery training, and (E) learned helplessness. Overall (1) the instructions for the games were clear; (2) the games were fun to play; and (3) the purpose of the games was clear.
Table 2. Summary of feedback on the games from 15 pilot users. Each game was rated on 3 parameters on a 5-point Likert scale, in which 2=strongly agree, 1=agree, 0=neutral, −1=disagree, −2=strongly disagree. The $P$ value corresponds to Wilcoxon signed rank test against 0.

<table>
<thead>
<tr>
<th></th>
<th>Values, mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automatic interpretation bias modification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>0.800 (1.014)</td>
<td>.02</td>
</tr>
<tr>
<td>Fun</td>
<td>0.867 (0.743)</td>
<td>.003</td>
</tr>
<tr>
<td>Purpose</td>
<td>0.733 (0.883)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Executive control training</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>0.800 (1.207)</td>
<td>.03</td>
</tr>
<tr>
<td>Fun</td>
<td>1.000 (0.926)</td>
<td>.004</td>
</tr>
<tr>
<td>Purpose</td>
<td>0.533 (0.743)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Negative attention bias modification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>1.533 (0.639)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fun</td>
<td>1.000 (1.195)</td>
<td>.01</td>
</tr>
<tr>
<td>Purpose</td>
<td>1.067 (0.961)</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Positive imagery training</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>1.533 (0.639)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fun</td>
<td>1.067 (1.032)</td>
<td>.005</td>
</tr>
<tr>
<td>Purpose</td>
<td>1.267 (0.798)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Learned helplessness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>1.133 (0.743)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fun</td>
<td>1.133 (0.639)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Purpose</td>
<td>1.133 (0.743)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

This paper describes the design of 4 serious games based on CBM paradigms and 1 serious game based on the learned helplessness theory. We expect that delivering the CBM paradigms in the form of serious games will increase the paradigms’ engagement and consequent effectiveness. Currently, cCBT is the predominant digital intervention modality used for depression. Despite some early attempts at combining cCBT and CBM [23,32,33], no well-evaluated and widely available software intervention offers both. One likely reason is that users are not motivated to use the CBM paradigms in their raw forms. Incorporating game-based CBM paradigms in digital interventions might serve the following 2 purposes: increase engagement with the overall interventions and serve as a complementary therapeutic approach to CBT. Overall, all the games received positive feedback on all 3 aspects—whether the instructions were clear, whether the game was fun to play, and whether the purpose of the game was clear (Figure 8). In addition, the learned helplessness game was able to achieve its purpose of explaining learned helplessness via experiential learning.

**Optimization of Game Paradigms**

Two of the CBM paradigms described in this paper, the negative attention bias training [92] and the positive imagery training [94], were previously converted into serious games. In particular, the negative attention bias training paradigm has inspired many serious games. However, only a few have been evaluated in research studies [62,107,108], whereas the rest are commercially available without evaluation. The Bias Bluster game [109] was developed based on the positive imagery training paradigm.

We made slight modifications to 3 of the CBM paradigms to make them suitable for use in the games while keeping their active ingredients intact. In the original automatic interpretation bias modification paradigm, a sentence was shown for 1000 milliseconds [21]. However, during our internal testing, we observed that 1000 milliseconds was not appropriate for all sentences and, in general, too short for nonnative English speakers. Therefore, we showed the sentence for a duration dependent upon the length of the sentence, computed by $100 + (TC / ARS)$ milliseconds, where $TC$ denotes the total number of characters in the sentence and $ARS$ denotes the average reading speed (estimated to be approximately 50 characters per 1000 ms).

In the original negative attention bias training paradigm, the images were presented in a $4 \times 4$ grid [92]. During the internal testing, we observed that a $4 \times 4$ grid on a smartphone screen makes the individual images very small and the emotions in the images indiscernible. Considering the higher penetration of smartphones than computers or tablets, we presented images in $2 \times 1$, $2 \times 2$, or $3 \times 2$ grids to ensure that the images and emotions could be seen easily.
In some of the previous implementations of the positive imagery training paradigms, the scenarios were presented in an auditory format and the positive resolutions were provided to the user by the program directly [25,94,95,97,100]. We presented the scenarios using a text-based format, similar to Mathews and Mackintosh [98], as the auditory format is less amenable to gameplay (e.g., an audio clip is slower to scan back and forth compared with text). Furthermore, we designed the game such that the user was required to come up with a positive resolution to the ambiguous scenario. This modification was inspired by the observation that the training was more effective in changing the mood if the interpretation was actively generated by the user [98].

Limitations and Future Directions
Feedback from the users in this study indicates that serious games based on CBM paradigms are acceptable. However, this study had a small sample size. There is a need for more studies with larger, more diverse, and clinical samples to test the real-world engagement and effectiveness of the games. In the future, games can be made more engaging using virtual reality.

Conclusions
Serious game–based mental health interventions are acceptable to users and have the potential to increase their engagement with digital mental health interventions. The designs we have provided can also be adapted into other interventions for many other disorders in which cognitive biases are involved [110]. We expect that the inclusion of serious games based on the CBM paradigms will help in improving the effectiveness and engagement of digital interventions.

Acknowledgments
The authors thank the members of the NG laboratory for testing the initial versions of the games and providing feedback on the content and user experience. In addition, the authors thank professor Arjun Ramakrishnan and Abin Thomas for help in recruiting participants for the pilot trial. They also thank Arun Shankar and Ranjeet Kumar for labeling the images used in the game for the negative attention bias modification. Face images used in the game for negative attention bias training were collected from multiple sources: White—Psychological Image Collection at Stirling [111]; African and Asian—tarrlab [112] (courtesy of Michael J Tarr, Carnegie Mellon University; funded by National Science Foundation award 0339122); and Indian—photographed by us. This work was supported by the Cognitive Science Research Initiative (CSRI) of the Department of Science and Technology (DST; grant DST/CSRI/2018/102). The NG laboratory is supported by the DST Science and Engineering Research Board Swarnajayanti Fellowship (grant SB/SJF/2021-22/04-C). The funding agencies had no role in the design or implementation of the study and in the interpretation of the results.

Data Availability
The deidentified data analyzed in the pilot study are available from the corresponding author upon reasonable request.

Authors’ Contributions
AG and NG conceptualized the project. AG searched the literature for cognitive bias modification paradigms and suggested the initial game ideas that were refined and finalized after discussion with NG and other authors. AG, JA, SB, BKS, LA, AA, ST, KM, SR, YA, and SG developed the games. AG and NG wrote the paper with inputs from all coauthors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Flowcharts describing game algorithms.

References


61. Oliveira E, Gonçalves MM, Caridade R, Rodrigues N. Ruminiation room: a serious game to deal with disturbing thoughts. JMIR Serious Games. URL: http://www.jmir.org/2014/7067104


111. Psychological Image Collection at Stirling (PICS). University of Stirling. URL: http://pics.stir.ac.uk/ [accessed 2023-04-23]

112. tarrlab. URL: https://sites.google.com/andrew.cmu.edu/tarrlab/ [accessed 2023-04-23]

Abbreviations

CBM: cognitive bias modification

CBT: cognitive behavioral therapy

cCBT: computerized cognitive behavioral therapy
Original Paper

Gamifying Cognitive Behavioral Therapy Techniques on Smartphones for Bangkok's Millennials With Depressive Symptoms: Interdisciplinary Game Development

Poe Sriwatanathamma¹, MFA; Veerawat Sirivesmas¹, PhD; Sone Simatrang¹, BFA; Nobonita Himani Bhowmik², MFA

¹Doctor of Philosophy Program in Design Arts (International Program), Faculty of Decorative Arts, Silpakorn University, Bangkok, Thailand
²Department of Game Design, New York University, New York, NY, United States

Corresponding Author:
Poe Sriwatanathamma, MFA
Doctor of Philosophy Program in Design Arts (International Program), Faculty of Decorative Arts
Silpakorn University
22 Borommaratchachonnani Rd
Khwaeng Taling Chan, Khet Taling Chan
Bangkok, 10170
Thailand
Phone: 66 917655317
Email: poesriwatan@gmail.com

Abstract

Background: There is serious concern over the annual increase in depressive symptoms among millennials in Bangkok, Thailand. Their daily routine revolves around the use of their smartphones for work and leisure. Although accessibility to mental health care is expanding, it cannot keep up with the demand for mental health treatment. Outside Thailand, multiple projects and studies have attempted to merge gamification mechanisms and cognitive behavioral therapy (CBT) to create mobile health intervention apps and serious games with positive feedback. This presents an opportunity to explore the same approach in Thailand.

Objective: This study investigated the development process of gamifying CBT techniques to support game mechanics in a visual narrative serious game, BlueLine. The primary target of this research is Bangkok’s millennials. In the game, players play as Blue, a Bangkok millennial who struggles to live through societal norms that influence his digital life and relationships. Through in-game scenarios, players will learn and understand how to lessen the impact of depressive symptoms via gamified interactions on their smartphones.

Methods: First, this paper follows each development step of solidifying BlueLine’s game structure by integrating the Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches (ABCDE) model and narrative in games. Second, the approach to select CBT and related therapeutic elements for gamification is based on suitability to the game structure. Throughout the process, CBT experts in Thailand have reviewed these scenarios. The approach forms the base of the player’s interactions throughout the scenarios in BlueLine, broken down into 4 types of gamified mechanisms: narrative, verbal interactions, physical interactions, and social media interactions.

Results: With the game structure based on the ABCDE model, BlueLine scenarios implement gamified mechanisms in conjunction with the following CBT and related therapeutic elements: behavioral activation, self-monitoring, interpersonal skills, positive psychology, relaxation and mindful activities, and problem-solving. In each scenario, players guide Blue to overcome his triggered dysfunctional beliefs. During this process, players can learn and understand how to lessen the impact of depressive symptoms through gamified interactions.

Conclusions: This paper presents the development process of gamifying CBT and related therapeutic techniques in BlueLine game scenarios. A scenario can harbor multiple techniques, including behavioral activation, self-monitoring, interpersonal skills, positive psychology, relaxation and mindful activities, and problem-solving. BlueLine’s game structure does not limit the fact that the same combination of CBT elements ties each gamified mechanism.

(JMIR Serious Games 2023;11:e41638) doi:10.2196/41638
Introduction

Background

Nowadays, most millennials in Bangkok, Thailand, have adopted a long-hours working culture, which allows little time to relax and unwind. Social media has become the platform of choice, whether it is used to socialize with others or to catch up on public health warnings, such as those issued during the COVID-19 pandemic [1]. From 2016 to 2020, the number of social media users in Thailand jumped from 32 million to 52 million, approximately 80% of the population, with an average of 2 hours and 55 minutes of web-based social exposure per day [2]. After a long period of social media use, a possible side effect is social comparison, such as seeking acceptance and confidence from external sources [3]. Thus, it plays an essential role in influencing one to develop depressive symptoms [4].

A total of 20,685 Bangkok millennials responded to Thailand’s Department of Mental Health survey in 2021 [5]. The risk of developing depressive symptoms in this group is frighteningly high at 48.97%, and, compared with provinces outside the capital, it is 4 times higher [6]. The survey was completed using the Patient Health Questionnaire (2-question version, 8-question version, and 9-question version) mental health survey system that is based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).

Since the outbreak of COVID-19, depressive symptoms have become more apparent among Bangkok’s millennials [5-7]. During the pandemic, there were constant actions to promote mental health care awareness by hospitals and schools to provide quality and sustainable mental health care services [8]. Nonetheless, exposure to help is still limited, as was the case especially during the COVID-19–related lockdown period [9,10]. More services must therefore explore other approaches to reach patients experiencing depressive symptoms.

Smartphones are becoming a unanimous part of our daily routine, creating a window of opportunity to increase access to mental health care [11]. This can be in the form of serious games (designed for purposes beyond entertainment), which are part of applied games; games based on cognitive behavioral therapy (CBT) constitute 1 of the 6 major categories of tested games for mental health [12]. This approach can benefit players with favorable emotional treatment through awareness, behavior changes, and relief of symptoms [13,14]. Mobile mental health intervention is affordable and can be used adjunctively with the patient’s existing treatment [15].

Integrating CBT Techniques Into Games

In recent years, studies have attempted to develop frameworks to create practical approaches to developing gamified apps. However, most still need to be fully established [16,17]. It is challenging to integrate CBT into different contexts of each game while ensuring that it achieves the intended experience and treatment goal for the players [18]. In this paper, the term intended experience refers to the expected gameplay experience of the player (ie, the quality of the gameplay interactions) [19]. It is plausible to integrate CBT techniques into games as multiple successful projects existing in different regions have shown (eg, a mobile health [mHealth] app for Māori [Indigenous New Zealand people] youth with emotional difficulties, Quest Te Whitianga) [20]. Night in the Woods (a winner of the British Academy of Film and Television Arts game award for narrative) is the story of Mae, who is dealing with depression and anxiety as she quietly returns to her hometown [21].

This paper breaks down the development process of a gamified intervention incorporating CBT techniques into BlueLine. The overall intended experience for BlueLine is for the player to live through scenarios as the main character where they are given choices and can perform actions that incorporate the following therapeutic components: behavioral activation, self-monitoring, interpersonal skills, positive psychology, relaxation and mindful activities, and problem-solving.

Methods

Overview

BlueLine is a visual narrative serious game where the player plays as Blue, a Bangkok millennial who struggles to live through societal norms that influence his virtual life and relationships. The game is inspired by Florence [22], in terms of the way the narrative expresses through visuals, and by We Should Talk [23], in terms of how the player’s verbal choices affect their emotional state and lead to certain consequences in the game’s world. This game primarily targets mobile platforms, and most Bangkok millennials use them, especially those who exhibit depressive symptoms. The game is also designed to be approachable to casual players. On release, BlueLine will be playable in both Thai and English to support the primary target group.

In the early stages of developing BlueLine, the focus was to create prototypes that could successfully integrate the gamification of CBT techniques into narrative design and harmoniously use visual design as the core gameplay. Our design process includes medical doctors and therapists who frequently care for Bangkok’s millennials. It allows us to circumvent the limited studies on the use of CBT in Thai games available at the time of this research. We will conduct a randomized controlled trial of the complete BlueLine prototype to assess its efficacy, acceptability, and usability with the support of Thailand’s Department of Mental Health.

First, we must understand the behavior of our primary target audience, the millennials of Bangkok. The behavior patterns of popular activities among this group establish statistics, especially the popular interactions in each social media application [24,25]. It also includes the time spent on each web-based activity. From a game design point of view, creating scenarios by researching and understanding the habits of the target audience can lead to better engagement and enable delivering of the intended experience [26]. Our design team has created these scenarios...
to be approachable even to people who do not have depressive symptoms. Through these CBT-gamified interactions, players can absorb information to become familiar with how to cope with depressive symptoms. They can fail safely without social consequences in real life. Each game scenario will also have its intended experience, aligning with its intended experience.

Second, we aimed to obtain insight from experts who have experience using CBT techniques to treat Bangkok’s millennials. For a better understanding of the game design of this project, we created short scenarios in the form of storyboards and gameplay mock-ups in the form of animation to help the experts to express their opinion on how the game will affect their patients and to suggest details that can improve the effectiveness of the CBT-infused interactions. These short scenarios are experimental prototypes that explore how we can merge CBT techniques and gamification elements. Successful prototypes that deliver the intended gameplay experience will be compiled into the game later.

The feedback from monthly meetings with the experts is critical in developing future iterations. Playtests were held within the team frequently, followed by group discussions, which were also immensely invaluable for our developers with CBT-related experience. They shared opinions on aspects that they feel are beneficial and need improvement. This technique of verbalizing thought processes in usability testing is called thinking aloud [27]. At this stage, the emphasis was on ensuring that the prototype was complying with the intended experience while incorporating placeholder art assets. This was a standard game development process to prove that the participative elements in the gameplay can be practical without having to rely heavily on aesthetics.

Selecting CBT Techniques as the Game Structure

CBT is an evidence-based treatment for psychiatric disorders, including depressive symptoms. This treatment involves attempts to change ways of thinking and behavior patterns through understanding negative thoughts, emotions, behaviors, and physiological reactions, thereby relieving their symptoms [28]. This helps individuals learn to be their own therapists [29]. For BlueLine, the game’s core structure implements the Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches (ABCDE) model. It is a simple mnemonic developed in the field of rational-emotive behavior therapy by Albert Ellis [30]. This model targets beliefs as a fundamental course of treatment. It guides the individual to break down the event and carefully examine the causes and effects so that they may respond to the situation in an improved manner if it recurs [31].

The core structure often integrates the ABCDE model in several CBT-based applications and serious games. One such application is Self-help, Integrated, and Gamified Mobile Phone Application, a gamified mHealth intervention, and it is used as an evidence-based theoretical structure for players to overcome their maladaptive beliefs [32], whereas RETHink is a clinically tested therapeutic game for college students to manage distress efficiently [33], and ReWIND is a role-playing game that allows players to overcome their generalized anxiety disorder through anxiety-causing encounters [34]. Similar to BlueLine, they all use the ABCDE model to challenge dysfunctional beliefs in a safe virtual environment for their audience. Following this model, BlueLine’s structure is divided into 5 steps (Textbox 1). These 5 steps can match well with the narrative in a visual narrative game structure, as shown in Textbox 2. We can resolve the narrative tension by disputing beliefs and practicing new approaches. In addition, this allows other CBT and related therapeutic techniques to work in conjunction and still perform efficiently, including behavioral activation [35], self-monitoring [36], interpersonal skills [37], positive psychology [38], relaxation and mindful activities [39], and problem-solving [40].

Textbox 1. The Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches (ABCDE) model as a game structure from a player’s point of view.

<table>
<thead>
<tr>
<th>A: activating event</th>
<th>B: beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Players identify their main character’s trigger during a distressing scenario in the game.</td>
<td>Players understand the underlying dysfunctional belief as the scenario plays out by observing how the main character reacts.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C: consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>The impacts of the main character’s beliefs influence physically within the game’s world (character actions and surroundings) and psychologically (emotions).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D: disputation of beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main character will attempt to challenge their dysfunctional beliefs through a series of player interactions in the game.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E: effective new approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main character manages to transform their initial dysfunctional beliefs. The player will witness positive changes in the main character’s mindset.</td>
</tr>
</tbody>
</table>
Textbox 2. A narrative scenario in the Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches (ABCDE) model.

**Activating event**
- Blue is forced to stay late at the office. Even after he gets home, he gets another call from work and has to stay up to work.

**Beliefs**
- He feels worthless. No matter how hard he works, the office is not happy.

**Consequences**
- He feels unappreciated, drained owing to lack of sleep, and loses interest in activities.

**Disputation of beliefs**
- He recognizes that the office is taking advantage of him and that he should not devalue his worth based on other people’s words.

**Effective new approach**
- He discusses with the management his plan to not work after business hours and starts to perform recreational activities instead.

BlueLine simulates scenarios in safe virtual environments; players have the comfort of time to identify dysfunctional beliefs before they can happen. This way, they can see the consequences of their choices through self-monitoring the game character. The game allows players to challenge dysfunctional beliefs through its main character, Blue.

Following this structure, we aimed to scale up BlueLine because we plan the gameplay length to be approximately 1 hour and 20 minutes. We aimed to manage players’ time on a single playthrough of BlueLine, thus lowering the risk of preoccupation with the game, which is one of the symptoms of internet gaming disorder in DSM-5, Text Revision [41]. Several factors are related to addictions, such as life satisfaction [42] and the types of games (on the web and offline) [43]. Statistically speaking, players who spend long weekly gaming hours (approximately 20 hours per week) are more likely to develop depressive and addiction symptoms [44].

BlueLine gameplay has 15 chapters, each lasting 5 minutes on average. The game will offer pause and resume features. Players do not have to finish a chapter in 1 session and have the option to resume from where they left off. Our players may not return to the game because a lack of motivation can be related to depressive symptoms [45]. This presents an opportunity for a reminder feature to encourage players to resume the game through timely notifications; for example, the game would deliver notifications during the evening but not late at night. This reminder feature also exists in internet-based CBT applications [46] and eHealth programs [47].

Storyboards are also developed during this process, with mock-ups created to test the composition of visual elements on a smartphone screen. While sketching the storyboards, many discussions arose about specific moments, leading to more inspiration to develop the narrative further. We can determine which CBT-infused action the main character could perform by seeing the sketched character in the scenario. We specifically designed the color palette and animation to ensure that players are not exposed to specific light patterns or flashing lights, which can lead to photosensitivity epilepsy or blackouts. Thus, storyboards guide the next step—the gamification of CBT techniques.

**Approach to Gamification of CBT**

The challenge is to gamify CBT techniques appropriately and fit them into each scenario. The chosen approach is based on case-by-case scenarios, with the end goal being that the interactions should be meaningful to the player. Meaningful play occurs when the player’s action has immediate importance and affects the larger context of the game [48]. We were able to accomplish this by basing the content of BlueLine on Bangkok’s millennials, including their daily routines, behaviors, and life experiences. This shapes the identity of Blue (the main character), who would face and overcome obstacles in the story.

**Gamification Elements**

We divide BlueLine’s CBT gamification elements into 4 main components: narrative, verbal interactions, physical interactions, and social media interactions. We acknowledge that existing studies discuss gamification elements in CBT-based applications, such as scores, levels, and tokenized rewards [49,50]. However, there are difficulties implementing them in BlueLine because it is not in the nature of a visual narrative game to have repetitive elements. Table 1 shows the core structure of gamification mechanisms, and other use cases will be discussed in the Results section.
Table 1. Gamified mechanisms in BlueLine.

<table>
<thead>
<tr>
<th>Gamification of CBT in mechanics and description</th>
<th>Use in BlueLine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Narrative</strong></td>
<td></td>
</tr>
<tr>
<td>Growth of Blue as a representation of Bangkok’s millennials</td>
<td>The story is told from the perspective of Blue and his millennial daily routine, lifestyle, and life choices. Through the ups and downs in the story, Blue struggles with the triggering of his dysfunctional beliefs and overcoming them. CBT techniques: self-monitoring, problem-solving, and interpersonal skills.</td>
</tr>
<tr>
<td><strong>Verbal interactions</strong></td>
<td></td>
</tr>
<tr>
<td>Engagement in conversation</td>
<td>Blue participates in conversations with NPCs throughout the game. Each NPC’s reply contains therapeutic components depending on the player’s verbal choices. CBT techniques: self-monitoring, interpersonal skills, and behavioral activation.</td>
</tr>
<tr>
<td><strong>Physical interactions</strong></td>
<td></td>
</tr>
<tr>
<td>Activities that require Blue’s physical action</td>
<td>Blue participates in multiple activities that require him to perform the task physically. Interactions range from basic actions such as walking to complicated tasks such as grocery shopping and giving a love letter. CBT techniques: relaxation and mindful activities and problem-solving.</td>
</tr>
<tr>
<td><strong>Social media interactions</strong></td>
<td></td>
</tr>
<tr>
<td>Scrolling through social media and likes</td>
<td>This feature mimics a social app where players can scroll through the social feed in the game’s world. The content in the feed is generated by design to match the intended experience of that scenario. It is not linked to any actual social apps. Generated posts are visual based to avoid negative trigger words. CBT technique: self-monitoring.</td>
</tr>
<tr>
<td>Photo mode</td>
<td>This feature is a version of the camera app on mobile phones. The scenes that players see through the lens are generated by design, inspired by actual real-life locations. CBT techniques: relaxation and mindful activities.</td>
</tr>
</tbody>
</table>

aCBT: cognitive behavioral therapy.

bNPC: nonplayer character.

**Results**

**Overview**

This section details several scenarios where each gamification mechanism is integrated with CBT and therapeutic techniques. There is a degree of flexibility in which specific scenarios can have multiple gamification mechanisms and therapeutic techniques, whereas the core structure includes the ABCDE model.

**Narrative**

Depressive symptoms can recur, and that is the takeaway from BlueLine, which shows through player interactions that CBT and therapeutic techniques can reduce the risk of relapse as long as it is maintained [51]. BlueLine will take players through multiple scenarios built under the expanded ABCDE structure (Tables 2 and 3). This version adds a chapter column to support the upscaled story. A DSM-5 column is added to assess Blue’s depressive symptoms at each stage of the story. This serves 2 purposes. First, it keeps the developers and consulting experts informed about exposure to depressive symptoms during the design and production stages. Second, it identifies moments in BlueLine where the game could encourage players to be open to seeking mental health care. This approach could circumvent the unwillingness of Asian communities to access help from mainstream services [52]. Throughout the game, visual keys are placed to subtly help the player gain confidence to reach out for help if needed; for example, a customized billboard of a real-life mental health care service will be placed far in the background but still legibly visible.
Table 2. BlueLine’s Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches (ABCDE) structure with a narrative framework and a Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) column.

<table>
<thead>
<tr>
<th>BlueLine: Chapter 9, scenario 3</th>
<th>DSM-5</th>
<th>ABCDE model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue and his partner have a minor quarrel over the photos he has taken.</td>
<td>A1b-A9</td>
<td>Activating event: Blue tries to apologize because the photos he took were not his best.</td>
</tr>
<tr>
<td>Line thinks that Blue is not trying hard enough and believes that he does not care.</td>
<td>—</td>
<td>Beliefs: Blue feels he has messed up. In the hope of being forgiven by Line, he keeps apologizing to her.</td>
</tr>
<tr>
<td>Line vents at Blue.</td>
<td>—</td>
<td>Consequences: Blue feels insignificant and worthless because she despises the photos he took.</td>
</tr>
<tr>
<td>The player stops Blue from panicking.</td>
<td>—</td>
<td>Disputations of belief: Blue stops apologizing (stays silent) so that Line can gather her thoughts.</td>
</tr>
<tr>
<td>Line realizes that Blue did try his best.</td>
<td>—</td>
<td>Effective new approach: Blue learns that by not panicking, he enables Line to understand his worth.</td>
</tr>
</tbody>
</table>

A1 to A9: assessment of depressive symptoms at each stage of the story.

bTo be filled in by the developers and consulting experts.

Table 3. An ABCDEc model breakdown of the conversation exchanged between Blue and Line in chapter 9 of BlueLine.

<table>
<thead>
<tr>
<th>Activating event</th>
<th>Beliefs</th>
<th>Consequences</th>
<th>Disputation of beliefs</th>
<th>Effective new approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative</td>
<td>Blue and Line are having breakfast after the sunrise photo shoot. The couple begins to quarrel.</td>
<td>Line thinks that Blue is not trying hard enough and believes that he does not care.</td>
<td>Line vents at Blue.</td>
<td>Blue cannot continue to apologize because the game prevents the player from saying “Sorry” by fading out the “Sorry” button.</td>
</tr>
<tr>
<td>Blue</td>
<td>Blue feels he has messed up. In the hope of being forgiven, he keeps apologizing (achieved by the player pressing the “Sorry” button).</td>
<td>Blue feels insignificant and worthless because she despises the photos.</td>
<td>Interpersonal skills (using silence): Blue uses silence so that Line can gather her thoughts.</td>
<td>Behavioral activation: Blue learns that sometimes a moment of silence can help others to regain their minds.</td>
</tr>
<tr>
<td></td>
<td>DSM-5(^3) symptoms:</td>
<td>A7: a sense of worthlessness</td>
<td>A7: a sense of worthlessness</td>
<td>A8: impaired concentration</td>
</tr>
<tr>
<td></td>
<td>• A7: a sense of worthlessness</td>
<td>• A8: impaired concentration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td>Line is angry because the photos taken by Blue are not good enough.</td>
<td>She thinks that the photos are bad because Blue does not care. She does not want to listen to any excuses or apologies.</td>
<td>Line feels unappreciated and blames Blue.</td>
<td>Interpersonal skills (giving recognition): Line realizes that her words have hurt Blue.</td>
</tr>
<tr>
<td></td>
<td>DSM-5 symptoms</td>
<td>• A2: lack of interest</td>
<td>A2: lack of interest</td>
<td>A8: impaired concentration</td>
</tr>
</tbody>
</table>

ABCDE: Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches.

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

\(^3\): A7 and so on: assessment of depressive symptoms at each stage of the story.

**Act 1: Refresh**

It is a new beginning for Blue, who has just returned to Thailand to start a new chapter in his life with a good office job and to be with family. The player will begin appreciating the sense of well-being radiated by a motivated office worker through self-monitoring Blue’s daily routine. Blue will encourage the player to complete relaxation and mindful activities throughout the day by rewarding them with positive outcomes. This is achieved by engaging in physical interactions that some might consider mundane tasks (eg, waking up at 5 AM to leave for work early to avoid the traffic). At the same time, Blue will show appreciation for a beautiful sunrise. Eventually, the same routine begins to gray out as his surroundings slowly pressure him, such as being asked to work overtime. This will subtly trigger the activating event of the ABCDE model. The player’s...
Interactions at this stage consist of following Blue’s attempts to find his way back to the well-being he once experienced.

**Act 2: Reaching Out**

After unsuccessful attempts at socializing, Blue meets Line, a girl who is an office worker just like him. They start to bond and become more and more close after each meeting via verbal interactions. Line’s characteristics serve multiple vital purposes in the game. She serves as a source for Blue’s development; at times, her actions and speech are similar to those of a therapist (to help reduce the anxiety the player might feel regarding their first therapeutic appointment). When Blue is feeling low, she helps him by performing therapeutic-infused actions (eg, using interpersonal skills). She also encourages Blue to participate in relaxation and mindful activities to reduce his negative emotions via positive psychology [38] (eg, visiting gardens and mountains). After some time, she will trigger a dysfunctional belief stemming from upward social comparison. This will allow Blue and the player to reciprocate her kindness and problem-solving proficiency using the therapeutic techniques they learn during the game.

**Act 3: Reconnect**

This act is set during the time frame of the 2020 COVID-19 pandemic in Thailand. Although the world changes, we want to present ways to handle mental health for our stakeholders. In the story, Blue’s parents ask him to return home during quarantine. He is hesitant to leave Line’s side because he only recently moved in with her, but she is supportive of his decision to go back home to take care of his parents. Slowly, however, distance begins to chip away at their bond. Depressive symptoms arise from activating events through social comparison in topics such as “It is safe to go on trips because there is no close contact if nobody goes on a trip,” even as the lockdown intensifies in Thailand. Ultimately, the two work together to dispute their different dysfunctional beliefs to save their relationship. The last problem-solving session involves using the related CBT techniques throughout the game to unfold the final problem. The game rewards the player with a fulfilling climax—narrative, gameplay, visuals, and sound at their peak—making the final interaction memorable for players so that they may recall the way they used behavioral activation to handle the situation were it to happen in real life.

**Verbal Interactions**

Interactions with other nonplayer characters (NPCs) are essential in this game. Through their voices and actions, the NPCs aim to provide therapeutic components and gamification strategies, such as reinforcement of successful behavior, and provide feedback on the player’s performance through prompts [53]. Players can make verbal decisions through the dialogue-crafting system (Figure 1).

The dialogue-crafting system is inspired by the sentence-spinner mechanic from the game We Should Talk. The game’s use of sentence crafting creates different reactions. It focuses on using a designated combination of verbal choices for players to string together. In BlueLine, we have set up 2 types of sentence crafters: the single-choice sentence crafter allows the player to change 1 part of the sentence, whereas the dual-choice sentence crafter allows 2 parts to be changed.

CBT components will be integrated based on the verbal combination put together by the player and the feedback provided through the NPC, Line (Blue’s partner); for example, as shown in Figure 2, this conversation aims to surround the player with a therapeutic interpersonal relationship to make them feel cared for in a supportive, nonjudgmental environment [54].

The gameplay screen shows Blue (the player) and Line (the NPC) at a café. Their conversation is shown in the bubbles positioned above their heads. At the bottom is the sentence crafter. The phrase “It does get crazy” is prefixed, as shown by the solid line, whereas the switchable portion is designated by a dotted line. Although the 2 choices are worded differently, they have the same underlying intention, which is to create positive emotions through evoking a sense of well-being [38]: for example, a player whose childhood was filled with joyous and noisy moments will be able to relate more to these good memories than players whose childhood was not as joyous and noisy. This follows the concept of piggybacking in game design, using what players already know instead of teaching them from scratch [55].

The example presented in Table 3 involves depressive symptoms experienced by both Blue and Line. Together, they review the photos Blue took. Line does not like the photos; hence, Blue tries to apologize to her in an attempt to calm her down. A “Sorry” chat bubble will appear when the player presses the “Sorry” button at the bottom of the screen. However, no matter how often the player presses the “Sorry” button, Line does not calm down. To resolve this issue, the player has to stop pressing the “Sorry” button. This would prevent Blue from apologizing and enable Line to gather her thoughts. Ultimately, Line realizes that her words hurt Blue; she appreciates this and says, “Sorry. Thank you for trying your best.” Throughout this scenario, self-monitoring, interpersonal skills, and behavioral activation are used.
Physical Interactions

In BlueLine, physical activities expose players to therapeutic components through actions. This chapter presents 2 examples: one concerns implementing art therapy as relaxation and mindful activities, and the other concerns problem-solving.

The first scenario occurs when Blue stops by to purchase a bouquet of flowers on Valentine’s Day. Instead of buying a premade bouquet, he makes one by hand (Figure 3). The goal is to teach the player that socializing is not the only means of engaging in relaxation and mindful activities. The process of art creation in art therapy also includes therapeutic elements such as self-exploration, which promotes an individual’s self-awareness in terms of seeing a more positive self-image [56]. Art creation is not tied to a single medium and can vary depending on the culture concerned; some examples are painting, music, and knitting [57,58].

To make the experience feel tangible, the player can decorate the bouquet using their fingers to drag and drop each flower. There is a minimum placement limit before the game can proceed to the next scenario, but the player can take their time until they want to move on. We consider not tying in reward-centric game elements, such as keeping score or points for tokenized rewards, and instead allowing the journey of the interactions itself to be the reward. From our experience, this helps player engagement to last longer.

There is a moment in act 3 when Blue goes to a supermarket during the 2020 COVID-19–related lockdown in Thailand. He picks up face masks, alcohol, tissue rolls, and other necessities, but only so many items are available on the shelves. The scenario at this stage of the game displays the risk of developing depressive symptoms during the COVID-19 pandemic in Thailand. As the population experiences more...
exposure to the disease and tunes in to information about COVID-19 on social media for at least 3 hours per day, there are more substantial effects on mental health in the form of anxiety and insomnia [7]. However, the process of problem-solving can be positively oriented. The solvable problem is set up for the player as a challenge and encourages their ability to resolve it effectively [40]. In Figure 4, the player will perform a swipe-up gesture with their finger to get both Blue (the player’s character) and Line (the NPC) to wear masks. The players can see the progress through visually animated feedback from the 2 characters.

Figure 3. BlueLine’s handmade flower bouquet: (A) before gameplay and (B) after gameplay.

Figure 4. Wearing face masks during the COVID-19 pandemic.
Social Media Interactions

According to the 2020 user behavior survey conducted by the Electronic Transactions Development Agency [59], Thai individuals spend 11 hours and 25 minutes on the internet daily, which is triple the amount compared with 2013. Instead of avoiding this topic, BlueLine uses it to visually compare a web-based lifestyle with a balanced one. Another objective is raising awareness of the players so that they can better understand the current lifestyles of Bangkok’s millennials. Social media interactions are represented as social media apps and a photo mode.

The social media app in BlueLine focuses on the social feed aspect; it mimics existing social media apps, such as the Facebook, Instagram, and Twitter apps. However, the content of the feed is generated by our design team; none of it is taken from actual sources. To date, numerous studies point toward the impact of social comparison on social media on personal well-being and depressive symptoms [3, 60, 61]. Regarding this aspect, BlueLine intends to show players through self-monitoring that Blue’s well-being improves after his fixation on social media reduces. This can be achieved through design-generated content (images, comments, and the number of likes) and controlling the time spent on social feed gameplay. Some posts do deal with social comparison to avoid players developing adverse mental health outcomes, such as a decline in psychological well-being [62], and making unfavorable social comparisons [63]; for example, a person may experience a lowering of self-esteem after they see someone showing off their material possessions in a post. In terms of duration, each social media gameplay session has been designed to last <1 minute to prevent players from engaging in heavy social media use and its associated adverse effects, such as self-harm behaviors [64] and a lowering of self-esteem [65].

As many social posts are accompanied by images, the interactive actions in the photo mode will be familiar ground for our target group. Players can use Blue’s phone to take photos of various things, such as food in restaurants and natural scenery (eg, mountains; Figure 5). All the locations presented in BlueLine’s photo mode are replicated from actual destinations in Thailand. This mode can catalyze players to venture outside and engage in relaxation and mindful activities (eg, visiting parks rather than shopping malls).

The player’s photos will be saved into a digital photo album that acts as a gratitude journal so that players can look back at such moments and share the photos on the internet. In this album, there will be positive prewritten notations made by Blue to help remind players of the events that occurred in these scenarios (Figure 6). Gratitude reinforcement can improve daily mood and lessen adverse effects (eg, depressive symptoms) [66]. However, multiple studies suggest that combined social comparison and materialism can negatively affect subjective well-being [67-69]. Maintaining a high level of gratefulness can lessen materialistic behaviors [70]. To increase accessibility to Blue’s gratitude journal, we have made it possible for players to access the album via 2 routes: the pause screen and the main menu.

Figure 5. BlueLine’s photo mode.
Discussion

Principal Findings

This study outlines the process of developing BlueLine, a gamified CBT intervention project on smartphones. BlueLine uses the ABCDE model as the game’s core structure to present the gamified CBT components (Table 1). The core structure integrates the narrative structure before the gameplay with visual elements. The early development is iterative and experimental, and it involves collaboration with CBT specialists in Thailand who have expertise and knowledge in caring for Bangkok’s millennials. Multiple playtests were set up with university students to help with the development process.

As a serious game, BlueLine falls into the visual narrative genre. Its gameplay lasts approximately 80 minutes, broken into 3 acts and 15 chapters. The game is split into short bursts to match a common habit of millennials—resorting to technology for a momentary escape from work—as well as their short attention span [71]. For the context to be meaningful to the target group, the daily routines, behaviors, and life experiences of Blue (the main character) reflect those of Bangkok’s millennials. In addition, interactions in BlueLine are inputted using finger gestures that mimic those involved in the daily use of smartphones.

Primarily, we have designed BlueLine to be used in adjunct with the patient’s primary treatment. It is not designed to tackle depressive symptoms entirely; instead, it focuses on enabling the player to understand the impact of depressive symptoms and learn how to lessen it. In most cases, our target group will not be playing the game while attending a session with their therapist; thus, BlueLine must be able to retain player engagement through gameplay. BlueLine’s gamified CBT mechanics (narrative, verbal interactions, physical interactions, and social media interactions) have been created to be flexible enough to fit into many narrative contexts so that the game encourages repetitions of the following CBT techniques: behavioral activation, self-monitoring, interpersonal skills, positive psychology, relaxation and mindful activities, and problem-solving.

Strengths

Experience plays a crucial part in this project. Our members have experience making small-scale visual narrative games such as Yumi’s Home [72], a first-person 2D narrative game about depression, familial obligation, and isolation. Furthermore, we have attempted to understand our target players and their behaviors in various ways. First, for 2 years, we were invested in researching the depressive symptoms experienced by Bangkok’s millennials as well as prereproduction prototypes. Second, the knowledge and experience of Thai CBT experts and the Department of Mental Health guided our design decisions in the development. There are 4 CBT experts, and each has >10 years of experience in psychotherapy; 2 experts are from the Department of Mental Health and a public psychiatric hospital, and 2 experts are faculty members of psychology departments at Thai universities. Thus, the project can scale up with confidence as we reach production.

To start with, we planned to localize BlueLine in both Thai and English. Those with Thai as their mother tongue can understand the nuances of the language and slang used by the characters, whereas English is meant for the increasingly international population in Bangkok. Players may switch the language at any given time via the options menu.

The story in BlueLine and the characters are based on the standard daily routines of Bangkok’s millennials as well as the real-life stories of development team members. We have identified specific experiences that resonate through the spectrum of millennials’ age range. The gameplay and visuals can be expressed on a deeper level in terms of attention to detail when non–game designers (artists and programmers) also participate in the design process [26].

Depressive symptoms can recur, and that is the takeaway from BlueLine, which shows players through interactions that CBT and therapeutic techniques can reduce the risk of relapse as long as it is maintained [51]. A game such as BlueLine is akin to an empathic listener. To empathize with another person is to understand their feelings, which is a level above mere acknowledgment. Serious games can make player more comfortable reaching out to the CBT-infused gameplay. Finally,
BlueLine can enable players to learn by guiding the main character to challenge and overcome their dysfunctional beliefs.

Limitations
Currently, this project does not accept external funding. The project’s scope is scaled to fit our current workforce, a team consisting of 5 members. Realistically, this is an ideal situation because we are able to maintain complete control over our design decisions. We intend all research and design processes to be free of external monetary influences (eg, investors). Once we complete the project, the app will be released in collaboration with the Department of Mental Health. We will consider a monetization model only after we have completed a randomized controlled trial of BlueLine with Bangkok’s millennials.

Studies have shown the potential of gamification for depression care, but more data are needed to establish the effectiveness of this approach [73,74]. Many factors come into play when considering what gamification needs to accomplish in a serious game; each player may have different depressive symptoms, and the game cannot personalize the experience for each individual. An improper gamified element can have an adverse impact on individuals with mental illnesses [75].

Regarding game addictions, some studies have addressed their adverse effects on an individual’s responsibilities (eg, education [76]) and personality (eg, adverse effects on self-esteem and behaviors [77]). In addition, BlueLine will be played on smartphones, which further influences users’ total screen time. During the scoping phase, we established that the gameplay length is limited to <2 hours and is divided into small chapters to avoid long sessions. In the game, we have designed the interactions to be distinct from nonserious games and promote positive activities such as art therapy.

The BlueLine gameplay length of 1 hour and 20 minutes can be a little problematic, due to factors such as poor concentration, and can result in a higher likelihood of game addiction [44]. When a game loop is long, players can keep playing it repeatedly, which may result in long hours of screen time and addiction. However, BlueLine’s pause and resume features allow players to take a break and return to the game later. In addition, it is within our scope to implement a feature to remind the player that they have been on the phone playing the game for too long. These features are used in internet-based CBT applications [46] and eHealth programs [47].

Of note, there are practical difficulties in developing mHealth apps in a higher middle-income country such as Thailand because of the lack of existing gamified mHealth apps for benchmarking. The lack of extensive statistical data on Bangkok’s millennials makes it challenging to design specific scenarios in the game. To ensure that our design decisions can lead to the intended experience, we rely on creating multiple iterations of each scenario.

Growing up in Thailand, some of us have noticed that the general public, that is, those who belong to an earlier generation compared with millennials, have negative stereotypical misconceptions of games; they view games as childish activities and believe that a grown-up should not waste time playing games. Moreover, even today, seeking and receiving mental health care is stigmatized in many countries [78] and can lead to social rejection of patients as well as self-rejection experiences [79]. This aspect can hinder the reach of BlueLine to the target group. However, it is worth the attempt, considering the benefits of making CBT accessible to a substantial proportion of Bangkok’s millennials. Thailand is one of many low- and middle-income countries looking at this technology as an opportunity [80].

Future Directions
In the future, we will collaborate with the Department of Mental Health to conduct a randomized controlled trial of BlueLine with Bangkok’s millennials to assess its efficacy, acceptability, and usability. On the basis of the results, further adjustments will be made before BlueLine and its service avenues are completed.

Conclusions
In conclusion, the ABCDE model can be the structure of visual narrative games. The gamification of CBT and related therapeutic techniques are interchangeable based on the scenario and its intended experience for the player. To reinforce the same gamified combinations to the player while avoiding monotony in the gameplay, narrative context and visual representation can be used to retain player engagement through the game experience. However, BlueLine’s development process is specifically focused on content related to Thai millennials; thus, its use may be less effective to other groups. BlueLine remains to be investigated in more detail regarding efficacy in adjunctive use with the patient’s existing treatment in Thailand.

Acknowledgments
The authors would like to thank Thai cognitive behavioral therapy experts Dr Varoth Chotpitayasunondh; Dutsadee Juengsiragulwit, MD; Dr Somboon Hatayiusuk, MD; and Dr Somboon Jarukasemthawee, as well as Thailand’s Department of Mental Health. The authors would also like to thank the members of the BlueLine team—Krittika Walaijai, Ratchanon Prapassornpong, and Kanitpong Tangtrongchit—for further stepping up their roles and the care they share for Blue and Line.

Conflicts of Interest
None declared.

References


23. We should talk homepage. We Should Talk. URL: https://www.weshouldtalkgame.com/ [accessed 2023-02-17]


72. Yumi's home. NYU Homepage. URL: https://gamecenter.nyu.edu/projects/yumis-home/ [accessed 2023-02-16]


74. Six SG, Byrne KA, Tibbet TP, Pericot-Valverde I. Examining the effectiveness of gamification in mental health apps for depression: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 29;8(11):e32199 [FREE Full text] [doi: 10.2196/32199]


Abbreviations

ABCDE: Activating Events, Beliefs, Consequences, Disputation of Beliefs and Effective New Approaches
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
mHealth: mobile health
NPC: nonplayer character

Edited by G Eysenbach; submitted 03.08.22; peer-reviewed by M Kapsetaki, H Alyami; comments to author 25.01.23; revised version received 01.03.23; accepted 31.03.23; published 12.05.23.

Please cite as:
Sriwatanathamma P, Sirivesmas V, Simatrang S, Bhowmik NH
Gamifying Cognitive Behavioral Therapy Techniques on Smartphones for Bangkok’s Millennials With Depressive Symptoms: Interdisciplinary Game Development
JMIR Serious Games 2023;11:e41638
URL: https://games.jmir.org/2023/1/e41638
doi:10.2196/41638
PMID:37171845

©Poe Sriwatanathamma, Veerawat Sirivesmas, Sone Simatrang, Nobonita Himani Bhowmik. Originally published in JMIR Serious Games (https://games.jmir.org), 12.05.2023. 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.
Health Promotion in Popular Web-Based Community Games Among Young People: Proposals, Recommendations, and Applications

Philipp Martin1,2,3,4*, PhD; Boris Chapoton5*, MSc; Aurélie Bourmaud1,3*, MD, PhD; Agnès Dumas1,6*, PhD; Joëlle Kivits1,7,8*, PhD; Clara Eyraud1*, MSc; Capucine Dubois9*, MSc; Corinne Alberti1,3,7*, PhD, Prof Dr; Enora Le Roux1,3,6*, PhD

1Université Paris Cité, Épidémiologie Clinique Evaluation Economique Appliquées aux Populations Vulnérables, Inserm, Paris, France
2Institut National d’Études Démographiques, UR14 – Sexual and Reproductive Health and Rights, Aubervilliers, France
3Assistance Publique des Hôpitaux de Paris, Hôpital Universitaire Robert Debré, Unité d’épidémiologie clinique, Inserm, Centre d’Investigation Clinique 1426, Paris, France
4GDID Santé, Paris, France
5Université Jean Monnet, Coactis UR 4161, Saint-Etienne, France
6Groupe de Recherche en Médecine et Santé de l’Adolescent, Paris, France
7Cité du Genre, Paris, France
8Réseau Francophone de Littératie en Santé, Paris, France
9Fil Santé Jeunes, Paris, France

* all authors contributed equally

Corresponding Author:
Philippe Martin, PhD
Université Paris Cité, Épidémiologie Clinique Evaluation Economique Appliquées aux Populations Vulnérables, Inserm
10 Avenue de Verdun
Paris, 75010
France
Phone: 33 0676606491
Email: philippe.martin@inserm.fr

Abstract

Background: Young people use digital technology on a daily basis and enjoy web-based games that promote social interactions among peers. These interactions in web-based communities can develop social knowledge and life skills. Intervening via existing web-based community games represents an innovative opportunity for health promotion interventions.

Objective: The aim of this study was to collect and describe players’ proposals for delivering health promotion through existing web-based community games among young people, elaborate on related recommendations adapted from a concrete experience of intervention research, and describe the application of these recommendations in new interventions.

Methods: We implemented a health promotion and prevention intervention via a web-based community game (Habbo; Sulake Oy). During the implementation of the intervention, we conducted an observational qualitative study on young people’s proposals via an intercept web-based focus group. We asked 22 young participants (3 groups in total) for their proposals about the best ways to carry out a health intervention in this context. First, using verbatim transcriptions of the players’ proposals, we conducted a qualitative thematic analysis. Second, we elaborated on recommendations for action development and implementation based on our experiences and work with a multidisciplinary consortium of experts. Third, we applied these recommendations in new interventions and described their application.

Results: A thematic analysis of the participants’ proposals revealed 3 main themes and 14 subthemes related to their proposals and process elements: the conditions for developing an attractive intervention within a game, the value of involving peers in developing the intervention, and the ways to mobilize and monitor gamers’ participation. These proposals emphasized the importance of interventions involving and moderating a small group of players in a playful manner but with professional aspects. We established 16 domains with 27 recommendations for preparing an intervention and implementing it in web-based games by
adopting the codes of game culture. The application of the recommendations showed their usefulness and that it was possible to make adapted and diverse interventions in the game.

Conclusions: Integrated health promotion interventions in existing web-based community games have the potential for promoting the health and well-being of young people. There is a need to incorporate specific key aspects of the games and gaming community recommendations, from conception to implementation, to maximize the relevance, acceptability, and feasibility of the interventions integrated in current digital practices.

Trial Registration: ClinicalTrials.gov NCT04888208; https://clinicaltrials.gov/ct2/show/NCT04888208

(JMIR Serious Games 2023;11:e39465) doi:10.2196/39465

KEYWORDS
health promotion; web-based community games; young people; interventional research; recommendations

Introduction

Background
Health promotion requires the creation of health-enhancing public policy with diverse and complementary approaches, particularly in the face of public health challenges (chronic diseases, social inequalities, and health systems issues) [1]. In this sense, health promotion is defined as the process of enabling people to increase their control over and improve their own health [2]. Areas for action include creating supportive environments, strengthening community actions, and developing personal skills. Health promotion interventions must then consider the importance of community and social dimensions, mobilizing social interactions as a lever for population health [3]. These social dimensions can reduce risk factors and increase the well-being of individuals.

Youth is a major period of empowerment and life evolution, with the formation of identity, social relationships, new responsibilities, and a professional future. It is also a period marked by health issues [4] that health promotion can address [2]. To do so, public health actors must understand the population’s concerns, habits, and culture to adapt their intervention strategies. Describing and understanding the characteristics of an identified group and studying its behaviors, expectations, and preferences enables the development of health actions with greater impact on the target audience and society [5,6].

Web-based games are attractive and widely used by young people, involving an environment in which players can explore, interact, and follow scenarios in association with characters and stories [7]. Games present relevant challenges for players who require different types of effort to solve tasks (ludic activities) or share experiences [7,8]. The “community” aspect of some web-based games means that, as in social network sites, players can connect with others, build web-based relationships, and find a “peer-to-peer” environment and support around common interests and shared game codes [9]. The community aspect can be centered on specific ages, lifestyles, backgrounds, or thematic characteristics such as interests, hobbies, pastimes, or games [9,10].

Some game-based health promotion interventions are promising, with positive results in terms of feasibility and short-term effects on knowledge and attitudes related to diet, physical activity, and lifestyle [7,11]. However, there is a need to address some key challenges in terms of the process and optimal conditions of implementation to understand how games can be the site of effective health promotion (intervention mechanisms) and research. First, researchers must describe the implementation means of such interventions. Poor conceptualization or bad implementation of an intervention can have a direct impact on its effectiveness [12]. To be effective, web-based games created for health promotion should include the following characteristics or features: feasibility, ease of use, storytelling, feedback, fun, and a proven focus on the health topic addressed [7]. In the case of web-based games that were not originally dedicated to health promotion and research, such data are missing.

The Scott project aimed to deliver a health promotion and disease prevention intervention via an existing web-based community game for young people. Indeed, our hypothesis was that intervening via popular games might enhance the acceptability (user-friendliness) and scope of our actions. The Scott intervention consisted of facilitating small-group health discussions in the game. We chose the Habbo game (developed by the video game company Sulake Oy)—a free, popular, and web-based game with social interactions in which it was possible to intervene.

Goal of This Study
On the basis of the Scott concrete experience, the aim of this study was to (1) analyze players’ proposals during the implementation of the intervention in the Habbo game, (2) elaborate on researchers’ recommendations for conducting a health promotion intervention in existing web-based community games among young people, and (3) describe the application of these recommendations in new interventions in the same Habbo game.

Methods

Overview
Our study is embedded in the Scott project, which is broadly a health promotion intervention research project aimed at implementing and evaluating an intervention in a web-based game. The first pilot study was conducted in 2020 [13]. The Scott project is registered on ClinicalTrials.gov (identifier NCT04888208).

During the implementation phase of the intervention, we conducted this qualitative study on the expectations and proposals of the participants in the first pilot study. Of all the
pilot study participants, a panel of 22 participants were present to participate in the study.

We conducted a qualitative study nested in the Scott intervention implementation based on (1) a questioning of players’ proposals for delivering health promotion within a web-based community game, (2) working groups with a multidisciplinary consortium of researchers for the elaboration of recommendations (based on our concrete experience), and (3) an application of these recommendations for new interventions in the same game. The stages and study processes are shown in Figure 1.

**Figure 1.** Stages and study process of the study.

### Scott Intervention

The Scott intervention was a health promotion and prevention intervention, initially a tobacco prevention initiative, integrated into a popular web-based community game and thus targeting adolescent and young adult game players (the game is officially permitted to be played by an audience aged ≥14 years). During this experiment, we investigated players’ proposals for pursuing health promotion within the game on any health themes.

### Theory

The entire theoretical framework of the intervention is described in Multimedia Appendix 1 [14-20]. The theoretical framework is based on the notions of social support, group motivation, and social learning [21], as applied to small web-based groups [14]. The mobilized theories are the theory of planned behavior [15], social learning theory [16], peer education [17] (with a link between peer education and social learning theory [18]), and the notion of small-group behavior change in web-based communities [14]. The main premise is that group discussion fosters the development of collective motivation and sufficient knowledge, norms, and skills to promote favorable health behaviors.

### Setting

We conducted the intervention in the French-speaking version of the web-based world Habbo. Habbo is an open-world web-based social network--type game with gamification elements. Habbo is a fun game with various activities with social interactions and different worlds in the game, allowing players to participate in discussion groups. The players can communicate with each other by sending limited character messages. The instant messages are limited to 50 characters. This results in many messages, a fast flow of messages, and multiple conversations in 1 space.

Ambassador players are players who are recognized for their attendance and good behavior. A team of moderators and a community manager ensure that the exchanges are fair. The developers of the Habbo game already allow external health actors to intervene in their game for health promotion interventions (not involving research) with the facilitation of the community manager and ambassador players. In the Scott project, we asked for the creation of a space in the web-based game (named a “room” inside the game) specifically for the project, with the appearance of a garden inside which a bus with a maximum capacity of 10 players (small group) was parked for the delivery of the intervention (Figure 2). A queue was set up to allow a limited number of participants to get access to the bus (10 players per session). We replicated the intervention framework of Fil Santé Jeunes (a French information service for young people [aged 12-25 years] in the field of health), which has been intervening in Habbo for several years. One of the research team members (CD) facilitated contact with Sulake and the transmission of needed experiences and skills. The relationship with Sulake was established entirely through the Habbo game’s community manager. She helped us prepare the intervention (practical arrangements, recruitment, the dissemination of information notes, the implementation of the intervention, and the distribution of measurement tools).
Moreover, she acted as a unique intermediary between the developer and researcher, especially with regard to the following: the temporality of the intervention, the content of the sessions, questionnaires, data collected, and feedback.

**Figure 2.** Garden and bus for the project (place of intervention).

---

**Facilitators**

A pair of facilitators (PM and ELR) were in charge of the animation and moderation of action. This pair was composed of 2 public health researchers, identified in the game as woman and man, to ensure a gender mix. They had a pseudonym integrating the term “health” to assert their professional position.

**Development and Preparation**

We developed the intervention content with a pluridisciplinary research team consisting of prevention workers, a psychologist (already leading a support group centered on adolescents’ issues in the Habbo game), the game’s community manager, sociology and social marketing professionals, and researchers involved in health promotion and intervention research. Two researchers (the facilitators) were previously trained by one of the regular health professionals operating in the game, the community manager of the game, and 3 experienced players to facilitate the intervention. They were trained to test the game’s commands (speak, exclude, silence, etc) before setting up the intervention.

**Involvement of Peer Players**

We worked with the ambassador players for the preparation of the first intervention (part 1) and application of the recommendations (part 3). These players were involved in the training of the facilitators (getting to grips with the features) and in the construction of the places of intervention in the game (Habbo players can create different spaces on the web).

In the game, players can collect badges (with no monetary value), which were used as incentives for this project. Badges were allocated to the participants before and after participation. To enhance novelty, we proposed the opportunity to create and provide original badges dedicated to this project (coconstruction vision) to some volunteer ambassadors.

**Contents and Processes**

We offered four 30-minute discussion sessions on 4 Thursdays in October 2020 at the proposed times of 11 AM, 3 PM, and 5 PM (12 sessions). We proposed different types of content concerning health issues and tobacco prevention (session 1: representations of tobacco, session 2: benefits of not smoking, session 3: resisting proposals from those around you, and session 4: helping others by making proposals for others’ intervention).

For the fourth and final session, we asked all 22 players (1 session per group; 3 groups=3 sessions) for their proposals and recommendations for pursuing health promotion actions in the game.
Participants
The participants were all volunteer French-speaking Habbo players. A preintervention survey was conducted in January 2020 to ensure that the average age of the given sample corresponded to that of a population of adolescents and young adults. Of the 83 responses, 75% (n=62) were from male participants, and the median age of the participants was 23 years, with a minimum age of 16 years and a maximum age of 32 years. The median age is in accordance with the World Health Organization’s definition of young people’s age range, which is 10 to 24 years [4].

Recruitment Process
We recruited study participants in October 2020. A week before the first session, we asked for a banner advertisement on the game’s log-in page with all the information about the study (objective, facilitator’s contacts, time, place, and subject). We also explained that not all players would be able to participate because of limited places and randomization. The participants were informed that they were participating in public health research. After inclusion, before the session integrating the qualitative study (session 4), the participants were informed that the purpose of this session was to collect their proposals for future interventions.

On the day of the first intervention, on the home page of the game, the participants had to fill in a questionnaire on baseline information (its completion constituted a criterion of eligibility for randomization). After the completion of the questionnaire, the participants were randomly provided either the intervention group badge or the control group badge. We planned 3 time slots for recruitment on the same day, at 11 AM, 3 PM, and 5 PM. For each time slot, a maximum of 10 participants received the intervention badge, which was valid for the 4 sessions (the day of randomization and the following 3 Thursdays). The control group, as well as the intervention group, had access to health information on preexisting websites posted in the project room.

Data Collected
Data were collected (1) from the qualitative study and the fourth session (focus group) on proposals, (2) from the working groups with a multidisciplinary consortium of researchers and feedback from the research facilitators, and (3) from observations of other interventions developed on the basis of the recommendations and applied by other facilitators.

1. Players’ Proposals for Intervention
During the last (fourth) session, the players from the intervention group were asked about the attractiveness of the intervention and their proposals for health promotion in the game. We asked them to insert health promotion and prevention interventions into the Habbo game but beyond the Scott intervention. We asked the following questions: “Would you be willing to help us do prevention activities in Habbo? What would you consider important? What would you want to implement?” We collected verbatim testimonies from messages that illustrated players’ proposals, elements of attractiveness, and the process of the intervention. No personal or identifying data (including pseudonyms) were collected for analysis. We collected the number of participants and number of messages sent during each session. Beyond these data, the facilitators collected personal feelings about the sessions, with field notes and points for animating and moderating interventions. We transcribed the participants’ testimonies and associated them with a participant code (S1, S2, ..., Si) to preserve anonymity.

2. Recommendations’ Elaboration
Iterative working group sessions were held among researchers (including facilitators) to share and analyze feedback from the intervention sessions and formulate recommendations for further interventions in the games. These data were in the form of raw notes and feedback reports.

3. Recommendations’ Application
Three authors (PM, ELR, and CE) applied these recommendations during the training of new intervention facilitators in the same game. PM, ELR, and CE were responsible for training and supervising these facilitators. The facilitators were (third-year) medical students carrying out a practical initiation into prevention as part of their training. Eight different interventions over 2 weeks (4 per week), with 3 facilitators per intervention (24 facilitators [third-year medical students]) and new health themes (sexual health, addiction, physical activity, food, and nutrition), were developed. They conceptualized these new interventions based on the recommendations on diverse health themes, such as addiction, nutrition, sexual health, and physical activity.

Analysis
We conducted a qualitative analysis based on (1) the chat transcripts of the players, (2) a description of the recommendations elaborated by the researchers’ working group, and (3) a description of the observation of the recommendations’ application.

1. Qualitative Analysis of “Players’ Proposals for Intervention”
We conducted a qualitative analysis of the participants’ proposals regarding elements related to health promotion interventions: context, audience expectations, intervention components, and implementation modalities [22].

We conducted a thematic analysis focusing on the proposals and process elements related to the acceptability, attractiveness, and implementation of interventions in a web-based community game specifically. We allowed the participants’ proposals to emerge during the sessions. The analysis was conducted based on raw proposals without predefined themes. The authors formulated proposals around the main themes and associated subthemes following these stages: initialization, construction, rectification, and finalization [23]. The initialization phase corresponded to the rereading of each session (thanks to the game’s conversation history). The researchers EL and PM took notes on the emerging themes and their impressions. These 2 researchers then carried out the construction phase by classifying and describing the different expectations and proposals of the participants through a first classification and organization of the themes. For the analysis, we accumulated several messages from the same participant to relate all their proposals (only if
they made several short messages in a row). The rectification phase corresponded to the distancing of the themes, notably by proposing them to the other researchers of the team, for adaptation and stabilization. This phase also allowed us to put forward the themes that brought real new scientific knowledge.

Finally, the finalization phase focused on organizing the themes according to the link or distance between them (organization of the “storyline” of the proposals).

2. Recommendations’ Elaboration

Following the intervention, the researchers provided feedback and developed recommendations for interventions specifically conducted in web-based community games. These recommendations were designed based on current recommendations for interventional research, on the experiences of the 2 researchers involved as facilitators (PM and ELR), and on the data from researchers who acted as observers during some sessions (BC, JK, and AD); discussed; adapted; classified; and validated by all the researchers associated with the project.

We analyzed domains and recommendations: the domains correspond to the important steps of preparation, conceptualization of the intervention, and realization of the intervention (implementation); the recommendations are concrete and detailed suggestions for how to meet the expectations for each domain.

3. Recommendations’ Application

We observed and formulated a description of the recommendation-based interventions developed, centered on components and implementation modalities, to illustrate the concrete application of each recommendation.

Ethics Approval

The Inserm Ethics Review Committee and Inserm Institutional Review Board (IRB00003888, IORG0003254, and FWA00005831) provided a favorable opinion for our study and approved its information circuit on July 7, 2020.

Results

Overview

A total of 22 players participated in the session for the collection of proposals. Proposals were collected from groups of players (3 groups of 9, 8, and 5 players, named S1 to S22 in the following results), and 1715 messages were posted (an average of 572 messages per session), of which 1155 (67%) were posted by the participants and 560 (33%) were posted by the facilitators and researchers (PM and ELR). The facilitators and researchers (PM and ELR) who conducted the focus groups considered that the third group interviewed, in addition to the other 2 groups, provided no new information. Thus, they considered that data saturation had been reached.

Players’ Proposals for Intervention

Thematic analysis of the participants’ proposals revealed 3 main themes and 14 subthemes related to their proposals and process elements for health promotion action within the web-based community game. Textbox 1 presents these themes.

Textbox 1. Key themes and subthemes identified in the session.

<table>
<thead>
<tr>
<th>Develop an attractive intervention in the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The need for varied and renewed health topics</td>
</tr>
<tr>
<td>• Entertainment components and formats for prevention</td>
</tr>
<tr>
<td>• Challenge as a motivation to participate</td>
</tr>
<tr>
<td>• An entertaining intervention that may be considered not serious enough</td>
</tr>
<tr>
<td>• The involvement of multiple stakeholders in the development and implementation of interventions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Involve peers in the development of the intervention in the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>• In favor of codeveloping the action</td>
</tr>
<tr>
<td>• Remaining anonymous for the fear of not being taken seriously</td>
</tr>
<tr>
<td>• Feeling of insufficient knowledge for delivering prevention themselves</td>
</tr>
<tr>
<td>• Experience as a skill for prevention</td>
</tr>
<tr>
<td>• Motivation and social support among peers as a lever for action</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mobilize and control gamer participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Limiting the number of participants to ensure the success of the action</td>
</tr>
<tr>
<td>• Trolls and essential moderation of exchanges</td>
</tr>
<tr>
<td>• In-game rewards as incentives</td>
</tr>
<tr>
<td>• Promote action outside the web-based game</td>
</tr>
</tbody>
</table>

We analyzed domains and recommendations: the domains correspond to the important steps of preparation, conceptualization of the intervention, and realization of the intervention (implementation); the recommendations are concrete and detailed suggestions for how to meet the expectations for each domain.
Develop an Attractive Intervention in the Game

All the interviewed participants expressed an interest in contributing to the development of a prevention intervention in a web-based game. Through 5 subthemes, the analysis highlighted the themes, expected components, and formats, as well as the different stakeholders involved in the design and implementation of the intervention.

The Need for Varied and Renewed Health Topics

Initially, the participants mentioned the importance of addressing a variety of health topics. One of the participants (S2) suggested using a polling system to guide the topics addressed. Two participants mentioned the importance of including topics around “anything harmful” (S11) and health consequences by highlighting “impact features” (S2). One of the participants suggested a period for renewing topics, with “new themes each week” (S11).

Entertainment Components and Formats for Prevention

The participants proposed several action components, mostly participatory and interactive components, such as quizzes, games, and discussion groups (S11 and S12), to structure interventions. One of the participants (S11) explained the need to create places in the web-based game dedicated to prevention in the game, such as theaters or apartments. Another participant (S12) highlighted the importance of using animation and games as levers of participation:

> There need to be animations and games. Those make it more engaging, and questionnaires are attractive too. You know, you can do everything and anything like that, with quizzes in the form of questionnaires. [S12]

One of the participants mentioned the possibility of asking questions through a question box system (S11 and S12). Another suggested a system in which participants can ask questions through forums (S6). The same participant (S6) mentioned the possibility of introducing a prevention campaign in the game in connection with official institutional programs. Some also considered links to existing resources, with access to information summaries.

Challenge as a Motivation to Participate

One of the participants proposed a call for poster production in the form of a challenge to encourage the participants’ motivation and production:

> ...a poster competition within the game, with the results displayed in room galleries of the game. [S8]

Another participant reminded us that there are already challenges in the game for players and that we only need to reuse this system for the intervention:

> Actually Habbo has been doing competitions for years, so that means just re-using the same idea, and suggesting that players make a poster about smoking. [S7]

In addition, one of the participants mentioned competitions with a prize to win:

> Later we could set up animation games with a monthly ranking. For example, the top 5 players could be given a badge. [S14]

An Entertaining Intervention That May Be Considered Not Serious Enough

When we asked whether an in-game prevention message is a good idea, we were told the following:

> I am not sure. Philou [the facilitator], maybe they won’t take it very seriously. [S3]

A participant (S1) warned that playful or entertaining elements may not lead players to consider the prevention message integrated into the action:

> I think [the over-entertaining aspect of the prevention game] may be completely counter-productive and undermine the message it is supposed to convey. [S1]

The Involvement of Multiple Stakeholders in the Development and Implementation of Interventions

When asked about the technical development of the action, some participants mentioned the help of the community manager, present and known to all. She is the one who grants permissions and accompanies the creation of new elements of the game. She would have the task of explaining the course of a multicomponent action (S12):

> So organize animators let’s say once a week in addition to a discussion, with that a monthly ranking, with ‘the question box’ for the follow-up...the news of the community manager to explain the whole system. [S12]

More generally, the participants suggested involving several actors in the game for intervention: the community manager, the health professionals, and the player ambassadors (S1, S12, and S14):

> I think one of you two [facilitators] should be involved, with a Habbo representative, a player, and with an event on the welcome page. That makes it a real thing that’s being organised. [S14]

As a result, the interviewed participants showed a willingness to take action with their fellow players.

Involve Peers in the Development of the Intervention in the Game

The thematic analysis highlighted the community dimensions within the game that can be levers of attraction to preventive action. These elements were organized into 5 subthemes.

In Favor of Codeveloping the Action

Most participants expressed a willingness to contribute to the development of preventive action within the game. Their role can be that of being a peer advocate or peer educator:

> If it can have an impact on other people and can help them, I would be very happy to take part...I would hope to be able to take part or be a facilitator, to interact with the players and in fact help them with topics. [S2]
Remain Anonymous for the Fear of Not Being Taken Seriously

If they were to be involved as peer educators in delivering actions, 3 participants (S8, S12, and S14) mentioned the need to be anonymous with a different pseudonym or avatar. Not being recognizable as a classical player would make other players listen to them more. One of the participants explained the need for a professional presence to be taken seriously:

*In fact the worry is that if there are no professionals with us the players won't take us so seriously, because we know each other.* [S12]

Indeed, some are well known by their game peers, and it would be difficult for them to be taken seriously:

*On the other hand if I appeared as a clone that would be fine, because I wouldn’t be recognized. So long as there is anonymity I have no problem.* [S14]

Feeling of Insufficient Knowledge for Delivering Prevention Themselves

Beyond the players’ self-projection as prevention actors, one of the participants emphasized his feeling of not being able to intervene on all health issues:

*I don’t feel I am capable of making a judgment about everything.* [S13]

He also pointed to his lack of experience in talking about certain topics:

*Personally I don’t think I am the best person [to do smoking prevention] since I have never smoked.* [S13]

Experience as a Skill for Prevention

Participants mentioned that experience then intervenes as a skill that can be mobilized by the players or professional facilitators (S1, S12, and S15):

*Yes, when it’s doing adapted prevention, which can also bring out your personal experience of the subject.* [S1]

*...with people who have been affected by the subjects being discussed perhaps.* [S15]

*I am not a good example although being a smoker myself I know all about the downsides.* [S12]

If mobilized for delivering prevention messages, a participant reported that his point of view would be more intelligible and impactful:

*Our personal experience will make it more impactful...I would have more to talk about and have more impact in what I say.* [S2]

Motivation and Social Support Among Peers as a Lever for Action

For a participant (S6), the involvement of peers could take place through a social support system between gaming peers through sponsorship, similar to existing support strategies outside web-based games:

*We could do it like our counterparts for alcohol, except that we are on Habbo...Yes, the units could be a series of badges and they can be used as tokens, like for alcoholism.* [S6]

This peer-to-peer social support can lead to web-based group challenges, such as the community challenge mentioned by one of the participants:

*One could also think about a community challenge with ranking results.* [S8]

However, another participant (S11) highlighted the difficulty of motivating and mobilizing other players through digital means, especially for prevention:

*Well, yes I would be happy to help someone stop smoking, but it’s tricky sitting with a PC, isn’t it?* [S11]

Mobilize and Control Gamer Participation

Beyond the involvement of web-based gaming peers as peer educators, the thematic analysis highlighted the ways in which players (the target audience) are engaged in prevention. In this regard, 4 subthemes were highlighted.

Limiting the Number of Participants to Ensure the Success of the Action

The participants expressed that it is more pleasant to play the game in small groups:

*Yes, it’s more pleasant in a small committee.* [S12]

Several participants emphasized the need to think about a prevention action that would mobilize small groups in the game (S2, S4, S5, and S12), especially to be able to debate properly (S2). One of the participants (S5) suggested limiting the number of participants to 10, always with 2 facilitators. It was also pointed out that several sessions are necessary (S2):

*With a big group (100) it’s likely to be more difficult because there will always be people to disturb others who really want to pay attention.* [S2]

*I loved the once-weekly format.* [S14]

Trolls and Essential Moderation of Exchanges

Many players also talked about disrupters within the game (trolls). One of the participants said that there will always be “those who distract people who want to listen” (S2). A participant indicated that moderation is mandatory for the proper conduct of the action:

*It’s obligatory, Philou, you were in luck here, but in the bus there are usually trolls.* [S1]

This participant considered ambassadors (players with good behavior in the game) as people who could support this moderation:

*with one or two ambassadors to moderate the players who come to provoke everyone in your INSERM sessions.* [S1]

In-Game Rewards as Incentives

The participants discussed how to reach players and enroll them in prevention action. Some participants (S1, S9, and S10) indicated the importance of rewards for engaging players:
To be very frank with you, if there aren’t any rewards to be had, and it’s just for prevention... you won’t get many followers. [S9]

The same participants pointed out that in the game, “badges” are collectibles and that many players want to acquire them:

Those who win a badge, players really like badges. [S6]

Then I also think there has to be a badge for each session depending on the theme (people will be more receptive). [S1]

However, 2 participants (S3 and S11) suggested that players who come only for the incentives (badges) would be less influenced by the prevention action:

Many of them will come just for the badge. [S3]

It’s true that if there are incentives to win there will be more players, but we don’t want them to come just for the badges. [S11]

One of the participant (S12) expressed that not implementing these incentives would be a way to select the most motivated individuals:

Those who are prepared to attend knowing there is nothing to win are sure to be the most sincerely motivated. [S12]

Promote Action Outside the Web-Based Game

A participant indicated the importance of taking action beyond the game, especially to reach more people:

To be honest, Habbo is not the way to mobilize the most people, you need to aim higher. [S7]

Another way to get players to participate in action would be to promote the action inside and outside the game among the Habbo community (on Twitter [Twitter, Inc] and fan websites). One of the participants suggested putting billboards about the prevention action in the game (S2). Another participant (S11) suggested that the community manager could advertise on the game’s home page and through alerts.

Yet another participant suggested recording the actions to replay them and reach more people:

in a public space in the game, with the Habbo viewing experience. Film it and replay it on fan sites [for fans outside the Habbo game itself]. [S1]

Another participant (S6) mentioned that the social networking site Twitter and game-related accounts can be used to advertise the action:

You can do ads on Twitter, on the Habbo news site, and in the bus you can work on themes around addictions like cigarettes. [S6]

Recommendations’ Elaboration and Application

Following the analysis of young people’s proposals and professionals’ experiences, we established 16 domains of 27 recommendations for health intervention in existing web-based community games (Multimedia Appendix 2). Each illustration describes the application of each recommendation during the development of new actions introduced in the game. Thus, Table S1 in Multimedia Appendix 2 integrates the fields, recommendations, applications of the recommendations in new interventions, and descriptions of specific new interventions. Multimedia Appendix 3 describes the interventions developed in response to these recommendations.

Discussion

Principal Findings

Our study highlights Habbo players’ proposals for developing interventions within the web-based game. These proposals emphasized the importance of serious interventions that need to involve and moderate a small group of people in a playful way. Using a reflexive approach, we also proposed recommendations for application in different interventions in general web-based community games. These recommendations highlight numerous suggestions for preparing and implementing an intervention, correctly incorporating it into web-based games, and adopting the codes of the web-based community game culture.

Meaning of the Findings

Our study showed that it is possible to work in fields that are not commonly explored and exploited by intervention research, with important recommendations for making good use of them. This study shows that it is possible to integrate the intervention into games that the target population already plays instead of creating a new game for a health intervention. We knew that it was possible to intervene in the game (following the Fil Santé Jeunes experience). Nevertheless, this study wanted to delve further by collecting concrete proposals from the players. Recommendations based on the Fil Santé Jeunes experience completed these proposals. This made it possible to bring together the contributions of various stakeholders (target population and professionals) via a participatory research approach [24,25].

The players indicated the importance of addressing a variety of health topics, that is, being responsive and diverse in health promotion responses. The study participants highlighted the importance of addressing many of the associated themes in health promotion action in a web-based game. More generally, a large number of web-based games are used for individual health, mainly for cognitive training or indirect health education [26], physical activity [27], managing chronic disease [28], preventing bullying [29] or preventing tobacco use [30]. Health actors have developed a very large number of games, and these games are generally novel interventions delivered to individuals, with different degrees of acceptability, attractiveness, and feasibility.

The participants highlighted the need for different types of involvement (the involvement of peers and facilitators) and ways to conceptualize and implement interventions for health. Web-based games can play a major role in the adoption of positive health behaviors, particularly because they promote players’ self-esteem, provide support and information, and facilitate interaction and communication [28] with a sense of community. The world of games can be a real support. For
example, avatars are important among young people as an extension of the self or as a way to further experiment with social relationships safely. A previous study has shown that among trans or gender-diverse youth, the use of an avatar facilitates the expression and consolidation of gender identity, with powerful mental health benefits [31].

We proposed several recommendations on how to act in existing games, including the consideration of participants at several levels: personalization, solicitation, animation, and moderation. Furthermore, in the field of mental health, Fleming et al [32] made the following recommendations for using web-based games for interventions [32]: a user-centered approach requiring exploration of group preferences, engaging and effective interventions with a description of game dynamics, cross-sectoral and international collaborations to seek out and identify the skills required for game development, and rapid testing and implementation to keep interventions engaging.

From our perspective, we highlighted new elements to intervene in already existing web-based games. Nevertheless, the literature still has a few examples of interventions integrated into existing games that are popular with the target audiences. Few studies have focused on the added value of using games rather than creating them, whether in response to health issues or economic issues related to development [33]. One of the studies, however, highlighted the use of existing commercial video games as therapies, studying these games as potential health interventions [34]. Colder Carras et al [34] noted that there is a lack of research on the use of commercial games for health and highlighted the need to design standardized protocols, study best practices to identify which games work on which health conditions, and model the complex and dense data of games.

In the future, it will be possible to understand the effectiveness of actions integrated into web-based games for health promotion by considering the influence of action implementation and processes in web-based community games.

**Implications for Decision Makers and Practitioners**

Decision makers and initiators of health promotion interventions must consider several approaches to intervene through web-based games beyond the community aspects described in this study. It is necessary to consider that it is possible to intervene by 3 means: fitting the intervention into an existing game, creating a game specifically to intervene, and reusing a game without modification.

For interventions in existing games, it is necessary to consider what is practically possible (the potential of action). It remains necessary to understand the game’s functioning to provide adapted interventions (functioning and adaptability). Some games do not allow for this kind of intervention because there is too strong a difference between the game objectives or gaming companies’ culture and health objectives (real feasibility). If an intervention is possible, it is necessary to acculturate the intervention to the codes of the game without modifying the game. This requires the training and adaptation of external stakeholders (acculturation). Moreover, intervening in a game is only possible if there is a good link between the developers and stakeholders. It is important to establish a strong and trusting relationship with developers, particularly to integrate interventions in games (relationship). This relationship should also be considered from a research perspective, as it involves considering the data to be analyzed beyond the intervention to be implemented. On the one hand, the researchers must present to developers the real issues of research and the importance of collecting data that can be analyzed according to sometimes rigorous and complex methodologies (such as randomized controlled trials). On the other hand, the developers must explain how the game works (in a way that does not have health promotion as its initial goal) to anticipate both the best potential for action and the way in which the research can be carried out (eg, acceptability and accessibility of participants’ data).

To develop a game for a health promotion intervention (similar to Hong et al [35]), it is essential to consider the levels of involvement needed. It is important to consider the need for technical skills to develop and maintain the game (technical skills). It is also important to consider the costs inherent to development and maintenance (costs). Moreover, it is essential to understand the uses of, expectations for, and attractions of the games (understanding), particularly the benefit of creating a new game (innovation). Developing a web-based game also takes time, with the risk of obsolescence of the game in the face of rapidly changing preferences (temporality and obsolescence). Creating a web-based game for health also raises the issue of reaching target audiences, who are already engaged with existing games.

According to the example set by some research [36], a third mode of intervention could be the replay of a web-based game as it is without inserting health content or even modifying the game’s functionality. Although this would be inexpensive, it would require a clear identification of the actual health benefits that an unmodified, commercial web-based game could provide. The integration of these games with offline health practices or services is also generally required. These 3 modes that we discussed are illustrated in Figure 3.

In general, it is important for stakeholders to know which strategy is the most relevant to the target population, considering its accessibility and expectations (realism). The health intervention may have objectives that allow incorporation into existing games or require the creation of new web-based spaces. In any case, it is important to consider the implementation stages, stakeholders’ roles (positions), and way to integrate populations. Incentives, elements of web-based games that can promote participation, can be used to integrate populations (incentives). In our experience, badges were attractive. The accomplishment of tasks and its valorization in game are sources of motivation for participation. For example, compensation with game’s money can be discussed.
Strengths, Limitations, and Future Directions

Our study is one of the few to investigate the potential of existing web-based community games for health promotion. It provides concrete methodological recommendations for the conceptualization and implementation of health actions in existing games. However, this study has limitations. We consulted the participants who were involved in our intervention, as the questions asked in this study occurred in the fourth session of the intervention, inducing a selection bias and not reaching the most distant players. In fact, we do not have information about the preferences of those who are more difficult to reach in the game. However, we collected the opinions of the most motivated players; they discussed the attractive elements of intervention for all players and how to develop and implement interventions for the most distant players. We developed our recommendations on 1 kind of web-based game (community based, with the possibility of development of elements in the game by the players).

We developed recommendations according to a structured methodology based on the experience and analysis of the project researchers. This methodology and the application of these recommendations in the development of new interventions demonstrated the internal validity of the recommendations. Recommendations must nevertheless be tested and validated in other games of this type, by other facilitators, with other intervention components, and for other health issues. Furthermore, the various functionalities offered by different web-based games require tempering the application of certain recommendations. Indeed, in our case, the game is mainly a place of web-based social interactions with gamification elements. Other web-based games might have features that are less community based. Therefore, our article focused on web-based community games with a socialization purpose. In addition, some of the recommendations may also be applicable outside existing games—not only to the development of new community health games but also to interventions using facilitation, communication, and participation techniques for health promotion.

Future research must continue to understand the processes involved in this type of intervention, depending on the audience and health issues being addressed. Indeed, interventions in web-based games can vary by topic (eg, differences in appeal between tobacco and mental health actions and differences in behavior change theories) and by participating audiences. It is still important to understand which population is using different popular web-based games to better understand whom we are reaching and what efforts need to be made. Knowing which population is attracted to health promotion activities remains important. Integrating target audiences (young people) into research processes remains essential in a community-based participatory research approach [24,25]. This will enable a better understanding of how to adapt and develop acceptable and attractive health promotion interventions in web-based games.

Finally, research must involve evaluations based on complex intervention research recommendations to measure the effect of health promotion interventions in games on health determinants targeted by researchers and actors (information, attitudes, skills, and behaviors), taking into account context effects. To do this, the methodologies must be designed in an interdisciplinary manner and be based on existing recommendations [12,37].

Conclusions

Health interventions integrated in existing web-based community games represent a potential for promoting individual health and well-being. It is necessary to adapt to the changing preferences and uses of target populations by thinking of...
innovative means for intervention research (levers, brakes, and costs). This also requires a permanent acculturation of the interveners and researchers to correctly intervene and analyze the games as a place and potential for health promotion among the target audiences. Involving these audiences is of considerable help in designing and implementing innovative health promotion interventions.

Acknowledgments
The authors thank the Institute for Public Health Research for funding this study; the members of the Scott project for their contributions to the project, especially Bruno Giraudreau, Laëtitia Minary, and Anita Burgun; and Fil Santé Jeunes for their hospitality and training in good practices for conducting actions in the Habbo world. They particularly thank the community manager (anonymity preserved) for her indispensable support and the common management of this intervention. Finally, they thank the Habbo players for their opinions and advice on how to improve health promotion actions in web-based games, their translator Duncan Fulton for proofreading the manuscript, and the members of the Inserm U1123 and CIC-EC1426 teams for their support in the conduct of this project.

Authors' Contributions
ELR and PM wrote and submitted the protocol for the approval of this study. All the authors participated in the development of the intervention content. ELR and PM implemented the Scott intervention and collected the study data. CE, ELR, and PM applied new intervention sessions based on the recommendations. All authors participated in the analysis, generation of recommendations, application of the recommendations, and writing and validation of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Description of the theoretical framework of the Scott intervention.
[DOCX File, 21 KB - games_v11i1e39465_app1.docx ]

Multimedia Appendix 2
Key recommendation elaboration and application for health promotion intervention in a web-based community game.
[DOCX File, 27 KB - games_v11i1e39465_app2.docx ]

Multimedia Appendix 3
Description of interventions created based on the recommendations.
[DOCX File, 755 KB - games_v11i1e39465_app3.docx ]

References


Impact of a Serious Game (#RedPingüiNO) to Reduce Facial Self-Touches and Prevent Exposure to Pathogens Transmitted via Hands: Quasi-Experimental Intervention

Marta Arévalo-Baeza¹, PhD; Alejandro Viuda-Serrano¹, PhD; Carmen Juan-Llamas², PhD; Pablo Sotoca-Orgaz¹; Iván Asín-Izquierdo³, PhD
¹Department of Education Sciences, Faculty of Medicine and Health Science, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain
²Department of Biodiversity, Ecology and Evolution (Biomathematics), Faculty of Biological Sciences, Universidad Complutense de Madrid, Madrid, Spain
³Physical Performance and Sports Research Center, Department of Sports and Computer Sciences, Faculty of Sport Sciences, Pablo de Olavide University, Seville, Spain
⁴Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain

Corresponding Author:
Marta Arévalo-Baeza, PhD
Department of Education Sciences
Faculty of Medicine and Health Science
Universidad de Alcalá
Campus Universitario C/ 19, Av. de Madrid, Km 33,600
Alcalá de Henares, Madrid, 28871
Spain
Phone: 34 651569136
Email: marta.arevalo@uah.es

Abstract

Background: After the COVID-19 pandemic, society has become more aware of the importance of some basic hygienic habits to avoid exposure to pathogens transmitted via hands. Given that a high frequency of touching mucous membranes can lead to a high risk of infection, it is essential to establish strategies to reduce this behavior as a preventive measure against contagion. This risk can be extrapolated to a multitude of health scenarios and transmission of many infectious diseases. #RedPingüiNO was designed as an intervention to prevent the transmission of SARS-CoV-2 and other pathogens through the reduction of facial self-touches by thoughtfully engaging participants in a serious game.

Objective: Facial self-touches should be understood as behaviors of limited control and awareness, used to regulate situations of cognitive and emotional demands, or as part of nonverbal communication. The objective of this study was to ensure that participants become aware of and reduce these behaviors through a game of self-perception.

Methods: The quasi-experimental intervention was applied to 103 healthy university students selected by convenience sampling and put into practice for 2 weeks, with 1 control group (n=24, 23.3%) and 2 experimental groups (experimental group with no additional social reinforcement interventions: n=36, 35%; experimental group with additional social reinforcement interventions: n=43, 41.7%). The objective was to improve knowledge and perception and reduce facial self-touches to prevent exposure to pathogens transmitted via hands not only in health multihazard scenarios but also in ordinary circumstances. The ad hoc instrument used to analyze the experience consisted of 43 items and was valid and reliable for the purpose of this study. The items were divided into 5 blocks extracted from the theoretical framework: sociological issues (1-5); hygiene habits (6-13); risk awareness (14-19); strategies for not touching the face (20-26); and questions after the intervention (27-42), designed as a postintervention tool assessing the game experience. Validation of the content was achieved through assessment by 12 expert referees. External validation was performed using a test-retest procedure, and reliability was verified using the Spearman correlation.

Results: The results of the ad hoc questionnaire, which were analyzed using the Wilcoxon signed-rank test and McNemar index to identify significant differences between test and retest for a 95% CI, showed that facial self-touches were reduced (item 20, $P<.001$; item 26, $P=.04$), and awareness of this spontaneous behavior and its triggers increased (item 15; $P=.007$). The results were reinforced by qualitative findings from the daily logs.
Conclusions: The intervention exhibited a greater effect from sharing the game, with interactions between people; however, in both cases, it was helpful in reducing facial self-touches. In summary, this game is suitable for reducing facial self-touches, and owing to its free availability and design, it can be adapted to various contexts.

(JMIR Serious Games 2023;11:e45600) doi:10.2196/45600

KEYWORDS
self-touching; face; serious game; health; pathogen transmission; hazard scenarios; body language

Introduction

Background

The global pandemic caused by COVID-19 forced the World Health Organization and health care authorities to establish a series of fundamental measures to prevent the spread of infection, which have modified social behaviors [1,2]; social distancing, use of masks, frequent handwashing, and avoiding touching the face. Although the available epidemiological data and studies on environmental transmission factors have not defined fomite-mediated transmission through hand contact with the mucous membranes of the facial T-zone (eyes, nose, and mouth) as the main route, the risk of infection via this route by SARS-CoV-2 and other pathogens remains, as revealed by Marquès and Domingo [3], and rapid and direct recontamination can occur despite handwashing [4,5]. The survival of SARS-CoV-2 and other pathogens on surfaces is not known with accuracy, but we can take the influenza virus as a point of reference, which can survive for up to 10 minutes on our hands [6]. We can touch the mucous membranes of our face every 2 minutes [7], with a self-touching frequency of up to 50.06 times per hour [8], especially on the mouth, followed by the nose and eyes [9]. The frequency of hand-to-face contact shows great diversity; appears to be context dependent; is influenced by cognitive, attention, and emotional demands; and exhibits substantial individual and locational differences. It is difficult to establish an average [5,10], but given that a high frequency of touching the mucous membranes can lead to a high risk of infection, it seems essential to establish strategies to reduce this behavior as a preventive measure against contagions [11-13]. More importantly, this risk can be extrapolated to a multitude of health scenarios and the transmission of other infectious diseases.

Spontaneous facial self-touches (sFST) are defined as touches on one’s own body, from the hands to the face, which include movements such as rubbing, scratching, stroking, or leaning on the hands, and which involve a proprioceptive activation of joints, muscles, and connective tissue, being able to activate skin sensitivity in a more or less conscious way, as pointed out by Spille et al [12]. These authors emphasize the importance of different triggers of sFST, distinguishing self-touches by self-regulation in situations where complex cognitive processes occur; by regulation of the intensity of emotions (anxiety, tension, clumsiness, uncertainty, discomfort, pleasure, and surprise); and finally, by the manifestation of emotions not expressed through words, which can be understood from nonverbal communication. Thus, we grouped the triggers of sFST by distinguishing compensation touches (as a manifestation of underlying emotions to adapt to a given context); warning touches (with prior warning such as itching, perspiration, tingling, hair brushing, imperfections, or lacrimation); and communicative gestures (without prior warning, with a triggering situation, and learned gestures, such as a hand on the chin while listening or waiting).

To reduce sFST, we can apply heuristics through specific interventions by avoiding errors of omission with signals and reminders; emphasizing the use of “if-then” plans; applying substitute or alternative behaviors; and developing an adequate and progressive reorganization of the brain [14], a process that requires observation and becoming aware of the trigger from a conditional strategy [15]. It could also be treated by consciously learning each touch action; differentiating it from its usual response chain; and replacing it with another action [16], such as using a handkerchief or touching another area of the face that is not a mucous membrane. In summary, the leading techniques for changing habits and behaviors seem to be repetition, the use of instructions and signals, action planning, the provision of instructions regarding how to carry out the behavior, and the establishment of behavioral objectives in the search for these self-control behaviors.

Objectives

This study was based on the serious game. This concept is understood as a carefully planned playful framework, the main purpose of which is to provide not only the entertainment itself but also an engaging context that involves people who are playing in the learning process [17-19]. Research has resorted to the use of serious games to address health care content, in the context of training, preventing disease, rehabilitating, or seeking to improve quality of life and well-being [20-22]. A serious game like the one analyzed in this study is a tool with complex development but simple intervention, which favors its use in various contexts, such as health institutions, hospitals, community centers, schools, or similar entities, and by the entire population, adapting the research procedures and design to the characteristics of the participants. This intervention does not require complex procedures, tools, techniques, or specialized software; therefore, it is accessible, efficient, and possibly helpful in reducing and perceiving the behavior of not touching the face. The #RedPingúINO methodological proposal is an intervention based on a serious game with the main objective of reducing facial self-touches as a preventive measure against exposure to pathogens transmitted via hands, such as SARS-CoV-2 and others. The foundation of its design is underpinned by the multidisciplinary theoretical framework, focused on a playable, engaging, and viral proposal that uses metaphors, humor, and a scoring system as drivers of the proposed game, using digital resources designed ad hoc that can be adapted to different contexts. Therefore, the objective
of this study was to analyze the effects of the #RedPingüINO intervention on the awareness and reduction of facial self-touches, especially in the T-zone, based on self-perception, given that a serious game seeks to positively impact the habits of participants.

**Methods**

**Research Design**

To design the research and the serious game, we have used Game-Based Intervention Reporting Guidelines given by Warsinsky et al [23]. Guidelines on conceptual issues, contributions of the study, related concepts, and definitions were taken into account; serious games were defined (Introduction section), the research stream was described (Methods section), the contribution of this study was clarified (Discussion section), and only core concepts based on the extant literature were used throughout the text. This study used a quasi-experimental design of involving a nonequivalent control group (CG) and a mixed approach. It is not possible to guarantee random assignment of the participants to the groups, but the educational center provides a predetermined distribution, ensuring limited homogeneity. This methodology is suitable for studying the potential causal effects of an intervention in open situations. The flowchart of the intervention is presented in Figure 1.

**Figure 1.** Flowchart of the intervention. CG: control group; EG: experimental group; EG-nR: experimental group with no additional social reinforcement interventions; EG-R: experimental group with additional social reinforcement interventions.
Serious Game Design

#RedPingüINO is a project based on a serious game that aims to reduce the spread of COVID-19 and other diseases through strategies to avoid the act of touching the face with one’s hands. The game uses penguins as protagonists because of their anatomy (they cannot touch their faces with their hands) and their social nature (collective and community behavior). The intervention platform has materials available in 8 languages. The game consists of associating sound onomatopoeias with behaviors: “No!” if the T-zone of the face is touched; “Ouch!” if another part of the face is touched; and “Bravo!” if the face is not touched at all. Descriptors and rankings are presented based on the daily frequency of self-touches (Figure 2). The rules of #RedPingüINO were published on the #RedPingüINO website [24].

Figure 2. Infographics and #RedPingüINO website.

Participants

A total of 125 students taking the Primary Education Teaching Degree course from a public university in Madrid participated in this research. The sample was selected by convenience sampling with 103 healthy participants who completed the study, comprising 74 (71.8%) women and 29 (28.2%) men; 22 (21.4%) dropped out during the intervention, as they did not complete 80% of the proposed activities. The average age was 22.76 (SD 2.78) years and gender differences in the sample were due to the specific characteristics of the chosen degree, with a larger female representation. The participants were distributed into three groups via the use of the Telegram app (Telegram Messenger Inc): (1) the CG (n=24); (2) the experimental group with no additional social reinforcement (EG-nR; n=36); and (3) the experimental group with additional social reinforcement (EG-R; n=43). Participation in the study was voluntary and anonymous and required informed consent.

Ethics Approval, Informed Consent, and Participation

This study was approved by the Ethics Committee of the Universidad de Alcalá (reference: CEI/HU/2020/46). The research complied with all national regulations and followed the tenets of the Declaration of Helsinki. The data of the participants were treated confidentially, and no personal information was accessed. Privacy was respected, and no personal information has been published. All the participants provided informed consent.

Instruments

An ad hoc self-administered questionnaire about facial self-touches and the game #RedPingüINO was administered to the participants using an easy survey platform before and after the intervention (the items in the instrument are listed in Multimedia Appendix 1). The instrument was made up of 43 items that were divided into 5 blocks and prepared from the following theoretical framework: (1) sociological issues (1-5), (2) hygiene habits (6-13), (3) risk awareness (14-19), (4) strategies for not touching the face (20-26), and (5) questions

https://games.jmir.org/2023/1/e45600
JMIR Serious Games 2023 | vol. 11 | e45600 | p.207
after the intervention (27–42). The fifth block was designed as a postintervention questionnaire and was administered only to the participants of the experimental groups (EGs). The items were closed ended, except for 1 open-ended question that collected considerations regarding the game (item 43).

The content validation of the questionnaire was conducted by expert judges [25]. A total of 12 judges (8 women and 4 men) assessed the content and form of the questionnaire. Their professional experience ranged from 11 to 35 years in scientific research related to the study, coming from various universities and heterogeneous fields of interest (body language, health, psychology, and didactics). The experts performed an overall assessment of the questionnaire and rated the degree of adequacy (relevance and wording). They were also asked about a set of aspects associated with the elaboration processes (appearance, clarity and length of the instructions, item format, and question order); data collection (completion time); and data interpretation (objectivity and absence of biases). The psychometric analysis of the content validity criteria of the instrument was performed using the V coefficient given by Aiken [26], considering the proportion of judges who express a positive assessment as a criterion for revising or eliminating items. An intervention pilot test and the registration were conducted by 4 teachers over a period of 5 days, during which they played the game and provided their assessments.

The participants of the EGs had to complete a daily self-compliance record of occurrence with 6 questions about the evolution of the intervention using the Microsoft Forms (Microsoft) tool and had to provide the following via Telegram: anonymous identification; frequency, prevention, and main reason for self-touches; motivation; and additional information.

In the web space created for the EG-R, social reinforcement actions were carried out, freely sharing, among the participants, experiences during the intervention, daily achievements, difficulties, and additional information.

**Procedures**

This research was coordinated and developed by faculty members and research staff who specialized in body language, with at least 10 years of experience. The questionnaires were administered by evaluators who were unaware of which group the participants belonged to, and the participants were unaware of the existence of other intervention groups.

The participants continued their daily routines throughout the intervention without supplementation of any other activity. The intervention was carried out for 2 weeks (November 23, 2020, to December 7, 2020) during the school term to foster the existence of a normalized context and to avoid the effect of holiday periods on the variation of routines. During this period, an attempt was made to modify the behavior of facial self-touches of the participants of the 2 EGs, EG-nR and EG-R, compared with that of the participants of the CG through the use of serious play. The reminders, sent via Telegram, occurred at the beginning of the day and around lunch time. At the end of the afternoon, a link to the registry was sent. The social reinforcement in the EG-R was in the form of praise, support, and approval, with tangible rewards and individual and collective reinforcement. The communication strategies used included cooperation, participation, interaction, and comparison.

**Data Analysis**

Statistical analyses were performed using SPSS package (version 25.0; IBM Corp). The types of variables in the questionnaire, namely, categorical and dichotomous, made it necessary to use nonparametric tests. To study the reliability of the questionnaire, different statistics were used to assess test-retest reliability. To analyze whether there were statistically significant differences between the answers given to the same items in the test and retest, the Wilcoxon signed-rank test was used, except for dichotomous variables, which were analyzed using the McNemar index. In addition, the degree of correlation between the test and retest items was determined using the Spearman rank correlation coefficient for nonparametric samples. To demonstrate the reliability of items 27 to 42, which were answered only by the participants of the EGs, the Cronbach α was used. Furthermore, to analyze the significant differences between the different groups, the following was used: for groups taken 2 by 2, the Mann-Whitney U test for 2 independent samples and (2) to compare the 3 groups, the Kruskal-Wallis test for 3 independent samples. Finally, content analysis was used to analyze the open-ended questions in the questionnaire [27].

**Results**

**Reliability and Validity of the Instrument**

The V coefficient by Aiken [26], which was used to determine validity, was interpreted using the score method [28], and it showed a very positive and adequate result for the use of the questionnaire. This procedure was developed based on the expert method, assuming a 95% CI (P=.05). The value of the test was V=0.99 for the adequacy of the items and V=0.96 for the wording, with minimum values of 0.88 and 0.87, respectively. Table 1 shows the reliability of the instrument for items 6 to 26. The test-retest values were adequate because they were .05; therefore, it was not necessary to remove any of the items. The Spearman rank test for nonparametric samples confirmed these results, because no correlations .05 were observed.

Items 12 and 13 of the yes or no responses regarding the frequency of facial self-touches with a mask and knowledge of the facial T-zone did not show statistically significant differences (P=.45 and P=.53, respectively). The participants of the CG showed a higher frequency of facial self-touches with a mask (13/24, 54%), and 75% (18/24) knew about the facial T-zone before filling out the questionnaire. The reliability of the instrument for items 27 to 42 related to the intervention was analyzed using the Cronbach α (.723). The Cronbach α coefficient showed a value .70, which is the usual acceptance limit [29]. In summary, the questionnaire was valid, reliable, and therefore suitable for use.
### Hygiene Habits, Risk Awareness, and Strategies to Avoid Touching the Face

The significant pre- and postintervention differences in the participants of the EGs in terms of hygiene habits, risk awareness, and strategies to avoid touching the face are shown in Table 2. As Table 1 shows that there are no statistically significant differences in the CG, this information is omitted from Table 2.

Hygiene habits showed significant differences ($P<.05$) at pre-and postintervention time points in several items, as shown in Table 2. The 2 EGs showed coincident positive significant differences in item 9, which is related to the disinfection of objects. Item 6, regarding the frequency of handwashing, establishes a significant positive difference in the EG-R and a negative difference in the EG-nR. In addition, the participants of the EG-nR improved the behavior of avoiding touching the facial T-zone (item 7). Items 12 (EG-R, $P>.99$; EG-nR, $P=.73$) and 13 (EG-nR; $P>.99$) did not show significant differences at pre- and postintervention time points, with the exception of the EG-R for item 13, knowledge of the facial T-zone ($P=.01$). The participants of the EGs showed a higher frequency of facial self-touches with a mask (45/79, 57%); and 77% (61/79) knew about the facial T-zone before filling out the questionnaire. Risk awareness establishes pre- and postintervention differences (EG-nR and EG-R) in item 19, showing a perception toward greater ease in avoiding the behavior of touching the face. Moreover, the EG-nR showed higher values in the perception of the risk of touching the face (item 15).

By contrast, regarding the strategies for not touching the face, as the main objective of this study, consistent significant differences were observed. In the pre- and postintervention analysis, both EGs showed a reduction in the perception and frequency of facial self-touches (items 20 and 26). However, the EG-R established a positive significant difference in the perception of the magnitude of this behavior (item 26); this result is possibly related to an increase in perceptive capacity. In addition, a clear improvement was observed in the perception of the possibility of reducing facial self-touches with training and the importance of awareness of the behavior among the participants of EG-nR (items 22 and 23) and in seeking a simple solution to avoid a risk behavior among the participants of EG-R (item 24). The significant differences between the groups are shown in Table 3.

Regarding hygiene habits, there were differences between the CG and the 2 EGs, and the comparison of the 3 groups in relation to the ability of avoiding touching the facial T-zone favored the EGs (item 7).

### Table 1. Reliability analysis of the questionnaire for the control group (n=24).

<table>
<thead>
<tr>
<th>Construct and item</th>
<th>Test values, mean (SD)</th>
<th>Retest values, mean (SD)</th>
<th>Wilcoxon sign-rank test or McNemar index ($P$ values)</th>
<th>Spearman rank coefficient ($P$ values)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hygiene habits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.63 (0.92)</td>
<td>2.71 (0.81)</td>
<td>.21</td>
<td>.80</td>
</tr>
<tr>
<td>7</td>
<td>2.33 (0.70)</td>
<td>2.54 (0.78)</td>
<td>.36</td>
<td>.46</td>
</tr>
<tr>
<td>8</td>
<td>2.92 (1.06)</td>
<td>2.79 (1.06)</td>
<td>.16</td>
<td>.54</td>
</tr>
<tr>
<td>9</td>
<td>2.67 (0.82)</td>
<td>2.58 (0.83)</td>
<td>.77</td>
<td>.46</td>
</tr>
<tr>
<td>10</td>
<td>2.83 (0.87)</td>
<td>2.88 (0.74)</td>
<td>.76</td>
<td>.70</td>
</tr>
<tr>
<td>11</td>
<td>3.92 (0.41)</td>
<td>3.83 (0.38)</td>
<td>.10</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Risk awareness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.54 (0.51)</td>
<td>3.42 (0.58)</td>
<td>.08</td>
<td>.81</td>
</tr>
<tr>
<td>15</td>
<td>3.63 (0.49)</td>
<td>3.54 (0.51)</td>
<td>.37</td>
<td>.41</td>
</tr>
<tr>
<td>16</td>
<td>3.58 (0.58)</td>
<td>3.58 (0.50)</td>
<td>&gt;.99</td>
<td>.23</td>
</tr>
<tr>
<td>17</td>
<td>3.42 (0.58)</td>
<td>3.54 (0.51)</td>
<td>.16</td>
<td>.54</td>
</tr>
<tr>
<td>18</td>
<td>3.58 (0.50)</td>
<td>3.58 (0.50)</td>
<td>.76</td>
<td>.33</td>
</tr>
<tr>
<td>19</td>
<td>2.17 (0.64)</td>
<td>2.38 (0.65)</td>
<td>.16</td>
<td>.73</td>
</tr>
<tr>
<td><strong>Strategies for not touching the face</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.83 (0.70)</td>
<td>2.04 (0.75)</td>
<td>.21</td>
<td>.65</td>
</tr>
<tr>
<td>21</td>
<td>3.54 (0.51)</td>
<td>3.54 (0.51)</td>
<td>.48</td>
<td>.46</td>
</tr>
<tr>
<td>22</td>
<td>3.54 (0.51)</td>
<td>3.38 (0.49)</td>
<td>.56</td>
<td>.26</td>
</tr>
<tr>
<td>23</td>
<td>3.71 (0.55)</td>
<td>3.46 (0.51)</td>
<td>.25</td>
<td>.38</td>
</tr>
<tr>
<td>24</td>
<td>2.96 (0.69)</td>
<td>2.75 (0.74)</td>
<td>.48</td>
<td>.47</td>
</tr>
<tr>
<td>25</td>
<td>3.42 (0.65)</td>
<td>3.29 (0.55)</td>
<td>.16</td>
<td>.60</td>
</tr>
<tr>
<td>26</td>
<td>2.46 (0.66)</td>
<td>2.50 (0.78)</td>
<td>.54</td>
<td>.33</td>
</tr>
</tbody>
</table>
With respect to risk awareness, positive significant differences were only observed between the CG and EG-nR, the CG and EG, and among the 3 groups in 2 items, 15 and 16, related to touching the face and handwashing. Finally, regarding the strategies for not touching the face, differences were observed between the CG and the 2 EGs, with higher rates observed in the EGs. These differences were observed for items 20, 24, and 25 in the EG-nR; for all items except 21 in the EG-nR; and for all except 21 and 23 in the full EG. In addition, differences were observed among the 3 groups for items 20 and 24. No significant differences were observed between the 2 EGs in any section.

### Table 2. Significant differences between pre- and postintervention measures in the experimental groups.

<table>
<thead>
<tr>
<th>Construct and item</th>
<th>Experimental group with no social reinforcement (n=36)</th>
<th>Experimental group with social reinforcement (n=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test values, mean (SD)</td>
<td>Retest values, mean (SD)</td>
</tr>
<tr>
<td><strong>Hygiene habits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.67 (1.06)</td>
<td>2.58 (1.02)</td>
</tr>
<tr>
<td>7</td>
<td>2.56 (0.65)</td>
<td>2.94 (0.79)</td>
</tr>
<tr>
<td>8</td>
<td>3.03 (0.97)</td>
<td>3.03 (0.81)</td>
</tr>
<tr>
<td>9</td>
<td>2.44 (0.81)</td>
<td>2.92 (0.73)</td>
</tr>
<tr>
<td>10</td>
<td>2.72 (0.85)</td>
<td>2.94 (0.63)</td>
</tr>
<tr>
<td>11</td>
<td>3.75 (0.50)</td>
<td>3.58 (0.69)</td>
</tr>
<tr>
<td><strong>Risk awareness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.56 (0.61)</td>
<td>3.64 (0.59)</td>
</tr>
<tr>
<td>15</td>
<td>3.67 (0.53)</td>
<td>3.86 (0.35)</td>
</tr>
<tr>
<td>16</td>
<td>3.81 (0.40)</td>
<td>3.86 (0.42)</td>
</tr>
<tr>
<td>17</td>
<td>3.58 (0.55)</td>
<td>3.72 (0.51)</td>
</tr>
<tr>
<td>18</td>
<td>3.75 (0.50)</td>
<td>3.78 (0.42)</td>
</tr>
<tr>
<td>19</td>
<td>2.11 (0.75)</td>
<td>2.47 (0.84)</td>
</tr>
<tr>
<td><strong>Strategies for not touching the face</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.11 (0.98)</td>
<td>1.44 (0.73)</td>
</tr>
<tr>
<td>21</td>
<td>3.61 (0.49)</td>
<td>3.72 (0.45)</td>
</tr>
<tr>
<td>22</td>
<td>3.42 (0.65)</td>
<td>3.64 (0.49)</td>
</tr>
<tr>
<td>23</td>
<td>3.50 (0.61)</td>
<td>3.72 (0.45)</td>
</tr>
<tr>
<td>24</td>
<td>3.08 (0.81)</td>
<td>3.28 (0.61)</td>
</tr>
<tr>
<td>25</td>
<td>3.36 (0.59)</td>
<td>3.61 (0.55)</td>
</tr>
<tr>
<td>26</td>
<td>2.58 (0.91)</td>
<td>2.00 (0.79)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistically significant differences (P<.05).
Table 3. Statistically significant differences among the control group (CG), experimental group with no additional social reinforcement (EG-nR), and experimental group with additional social reinforcement (EG-R).

<table>
<thead>
<tr>
<th>Construct and item</th>
<th>Mann-Whitney U test for CG and EG-R (P value)</th>
<th>Mann-Whitney U test for CG and EG-nR (P value)</th>
<th>Mann-Whitney U test for EG-nR and EG-R (P value)</th>
<th>Mann-Whitney U test for CG and EG (P value)</th>
<th>Kruskal-Wallis test for CG, EG-nR, and EG-R (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hygiene habits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.69</td>
<td>.66</td>
<td>.38</td>
<td>&gt;.99</td>
<td>.67</td>
</tr>
<tr>
<td>7</td>
<td>.01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.047&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.71</td>
<td>.01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>8</td>
<td>.80</td>
<td>.43</td>
<td>.16</td>
<td>.80</td>
<td>.38</td>
</tr>
<tr>
<td>9</td>
<td>.08</td>
<td>.10</td>
<td>.59</td>
<td>.06</td>
<td>.15</td>
</tr>
<tr>
<td>10</td>
<td>.12</td>
<td>.81</td>
<td>.11</td>
<td>.29</td>
<td>.17</td>
</tr>
<tr>
<td>11</td>
<td>.81</td>
<td>.18</td>
<td>.17</td>
<td>.39</td>
<td>.26</td>
</tr>
<tr>
<td>12</td>
<td>.48</td>
<td>.34</td>
<td>.76</td>
<td>.36</td>
<td>.63</td>
</tr>
<tr>
<td>13</td>
<td>.57</td>
<td>.54</td>
<td>.94</td>
<td>.50</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Risk awareness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.48</td>
<td>.10</td>
<td>.26</td>
<td>.20</td>
<td>.24</td>
</tr>
<tr>
<td>15</td>
<td>.054</td>
<td>.007&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.34</td>
<td>.007&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.02&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>16</td>
<td>.14</td>
<td>.009&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.17</td>
<td>.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.03&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>17</td>
<td>.17</td>
<td>.12</td>
<td>.78</td>
<td>.10</td>
<td>.24</td>
</tr>
<tr>
<td>18</td>
<td>.49</td>
<td>.11</td>
<td>.34</td>
<td>.21</td>
<td>.29</td>
</tr>
<tr>
<td>19</td>
<td>.88</td>
<td>.61</td>
<td>.75</td>
<td>.73</td>
<td>.89</td>
</tr>
<tr>
<td><strong>Strategies for not touching the face</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>.003&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.61</td>
<td>&lt;.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.002&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>21</td>
<td>.49</td>
<td>.15</td>
<td>.43</td>
<td>.25</td>
<td>.37</td>
</tr>
<tr>
<td>22</td>
<td>.06</td>
<td>.05&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.97</td>
<td>.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.09</td>
</tr>
<tr>
<td>23</td>
<td>.62</td>
<td>.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.08</td>
<td>.18</td>
<td>.09</td>
</tr>
<tr>
<td>24</td>
<td>.008&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.003&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.98</td>
<td>.002&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.008&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>25</td>
<td>.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.77</td>
<td>.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.051</td>
</tr>
<tr>
<td>26</td>
<td>.11</td>
<td>.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.40</td>
<td>.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.08</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistically significant differences (P<.05).

#RedPingüiNO Intervention

The intervention showed that 94% (74/79; EG-nR: 34/36, 94% and EG-R: 40/43, 93%) of the participants in both EGs achieved greater awareness of the times they touched their faces (item 27), and 95% (75/79; EG-nR: 34/36, 94% and EG-R: 41/43, 95%) of the participants perceived that they had managed to reduce the number of facial self-touches (item 28). There were no significant differences between both groups (P=.51 and P=.86, respectively). The intervention improved the awareness of the reasons why the participants touched their face in each situation (item 42), with a rate of 3.53 (SD 0.60) (mean 3.47, SD 0.56, for the EG-nR and mean 3.58, SD 0.63, for EG-R) on a 4-point Likert scale (1=totally disagree and 4=totally agree). Table 4 shows the results of the strategies, materials, and evaluation of the #RedPingüiNO intervention.
### Strategies and Materials

On a 4-point Likert scale, where 1 was “Little” and 4 was “A lot,” the intervention groups showed high rates in the perception of the different materials, highlighting recording (31) and video (item 36) over infographics (item 32). The strategies in the game of saying “No!” (item 29) and saying “Bravo!” (item 30) had high and similar results in both the groups. The follow-up strategy (item 33) was assessed by the participants of the EG-R at a higher rate than those of the EG-nR, without it being significant ($P = .07$). The Telegram strategies with high rates in both groups also did not establish significant differences between the 2 EGs (items 34 and 35).

### Evaluation of the Intervention

The intervention was assessed by the participants of the 2 EGs as engaging (item 37), helpful (item 38), and motivating (item 39), with high rates; however, the EG-R showed a higher rate than the EG-nR in the perception of the appeal of the intervention ($P = .045$). In both groups, the duration was perceived as sufficient, with high means (item 40) and a higher rate in the EG-R than in the EG-nR ($P = .001$). In addition, the serious game proposed as an intervention strategy was perceived as suitable for use within the school context (item 41), with high rates in both groups. Therefore, although the results are satisfactory in both groups, social reinforcement seems to have an impact on the perception of the intervention with higher means in the EG-R in all cases, regardless of whether they establish significant differences. This highlights the perception of the intervention as an engaging way to train awareness to not touch the face long enough to achieve changes in this behavior.

### Development of the Daily Intervention Log

In the analysis of the development of the daily log, a significant positive difference is observed in relation to the times that the participants of both EGs consciously stopped their hand before touching their face with a “Bravo!” (EG, $P < .001$; EG-nR, $P = .01$; and EG-R, $P < .001$), as well as a significant negative difference in relation to the frequency in which the facial T-zone was touched, and they realized it with a “No!” (EG, $P < .001$; EG-nR, $P = .049$; and EG-R, $P < .001$), possibly due to the significant reduction in facial self-touches. These data must be understood as the perception of a decrease in the frequency of facial self-touches. The reasons for facial self-touches were not modified among participants throughout the intervention. The participants of the EG-nR and EG-R established the reasons for touching their face as itching, lacrimation, and tingling (EG-nR: 17/36, 47%, preintervention and 16/36, 44%, postintervention; EG-R: 26/43, 61%, preintervention and 21/43, 49%, postintervention); restlessness, discomfort, and boredom (EG-nR: 13/36, 36%, preintervention and 12/36, 33%, postintervention; EG-R: 10/43, 23%, preintervention and 9/43, 21%, postintervention); and learned gestures (EG-nR: 6/36, 17%, preintervention and 8/36, 22%, postintervention; EG-R: 7/43, 16%, preintervention and 13/43, 30%, postintervention). The motivation perceived by the participants throughout the intervention showed a significant improvement, on a 5-point Likert scale, in the EG-nR (3.47, SD 0.87 at preintervention and 4.06, SD 1.09 at postintervention) and in the full EG (3.59, SD 1.25 at preintervention and 3.93, SD 1.16 at postintervention). This difference is not established in the EG-R (3.65, SD 1.40 at preintervention and 3.86, SD 1.21 at postintervention), owing to the perceived high motivation at the start of the intervention.

---

### Table 4. Significant differences between experimental groups (EGs).

<table>
<thead>
<tr>
<th>Item</th>
<th>EG, mean (SD)</th>
<th>EG-nR(^a), mean (SD)</th>
<th>EG-R(^b), mean (SD)</th>
<th>Mann-Whitney U test score for EG-nR and EG-R ($P$ value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>3.13 (1.15)</td>
<td>3.33 (1.07)</td>
<td>2.95 (1.19)</td>
<td>.09</td>
</tr>
<tr>
<td>30</td>
<td>2.92 (1.2)</td>
<td>2.78 (1.24)</td>
<td>3.05 (1.15)</td>
<td>.36</td>
</tr>
<tr>
<td>31</td>
<td>3.05 (1.1)</td>
<td>3.11 (1.21)</td>
<td>3.00 (1.00)</td>
<td>.26</td>
</tr>
<tr>
<td>32</td>
<td>2.53 (1.34)</td>
<td>2.33 (1.33)</td>
<td>2.70 (1.34)</td>
<td>.25</td>
</tr>
<tr>
<td>33</td>
<td>2.77 (1.25)</td>
<td>2.44 (1.34)</td>
<td>3.05 (1.11)</td>
<td>.07</td>
</tr>
<tr>
<td>34</td>
<td>3.05 (1.02)</td>
<td>2.83 (1.13)</td>
<td>3.23 (0.90)</td>
<td>.13</td>
</tr>
<tr>
<td>35</td>
<td>3.04 (0.85)</td>
<td>3.08 (0.87)</td>
<td>3.00 (0.85)</td>
<td>.60</td>
</tr>
<tr>
<td>36</td>
<td>2.94 (1.23)</td>
<td>2.72 (1.30)</td>
<td>3.12 (1.16)</td>
<td>.18</td>
</tr>
<tr>
<td>37</td>
<td>3.57 (0.61)</td>
<td>3.44 (0.61)</td>
<td>3.67 (0.61)</td>
<td>.045(^c)</td>
</tr>
<tr>
<td>38</td>
<td>3.48 (0.64)</td>
<td>3.39 (0.60)</td>
<td>3.56 (0.67)</td>
<td>.13</td>
</tr>
<tr>
<td>39</td>
<td>3.49 (0.62)</td>
<td>3.39 (0.60)</td>
<td>3.58 (0.63)</td>
<td>.10</td>
</tr>
<tr>
<td>40</td>
<td>3.16 (0.72)</td>
<td>2.89 (0.71)</td>
<td>3.40 (0.66)</td>
<td>.001(^c)</td>
</tr>
<tr>
<td>41</td>
<td>3.63 (0.58)</td>
<td>3.53 (0.56)</td>
<td>3.72 (0.59)</td>
<td>.054</td>
</tr>
<tr>
<td>42</td>
<td>3.53 (0.60)</td>
<td>3.47 (0.56)</td>
<td>3.58 (0.63)</td>
<td>.27</td>
</tr>
</tbody>
</table>

\(^a\)EG-nR: EG with no additional social reinforcement.
\(^b\)EG-R: EG with additional social reinforcement.
\(^c\)Statistically significant differences ($P < .05$).
**Discussion**

**Principal Findings**

The main objective of the #RedPingüiNO intervention was to reduce facial self-touches as a preventive measure against SARS-CoV-2 and other pathogens in various health contexts. The newly constructed tools developed for analyzing the intervention established adequate validity and reliability. After the 2-week intervention, the participants perceived improvement in their hygiene habits strategies, such as handwashing; disinfection of objects; the use of a handkerchief; and the action of avoiding touching the eyes, nose, and mouth. After the intervention, the perception of the risk that facial self-touches entail increased within the EGs, but at the same time, touching the face was noticed as an easier behavior to avoid. The participants in the 2 EGs perceived that they had reduced the number of facial self-touches following the intervention. Participants’ perception of the specific number of times they touched their faces also increased. They became more aware of the reasons why they touched their faces on a daily basis, the frequency of this behavior, and the importance of this awareness for self-regulation and avoiding self-touches. #RedPingüiNO was considered an appropriate and fun experience. More than half of the people who participated rated the game with qualifiers like “satisfactory,” “rewarding,” “positive,” “fun,” “useful,” or “surprising”; moreover, during the intervention, motivation increased as they perceived the achievement.

**Comparison With Prior Work**

This study seems to reinforce that changes in health behaviors are possible with appropriate tools. This need is based on the idea that the most effective steps to reduce infection and death rates during the pandemic have been based on changes in individual behavior [2]. After the intervention, touching the face was noticed as an easier behavior to avoid, which coincides with the approach of Heinicke et al [30]. These authors indicate that reducing or stopping facial touching is a habitual and prevalent behavior that is unlikely to decrease unless we train with specific strategies for self-regulation, which implies awareness and the establishment of effective responses. Facial self-touches seem to increase in frequency and duration in emotionally or cognitively challenging social situations. This may be due to the attention states, working memory processes, and emotion-regulating functions [12] associated with such situations, which are developed and maintained based on the body language used in interactions with oneself and other people. Itching, lacrimation, and tingling were the most noted triggers of this behavior among the participants who subsequently experienced restlessness, discomfort, and boredom; the participants were somewhat less alert to the learned gestures.

The increased perception of a reduction in the number of facial self-touches seems to respond to the fact that during serious play, there was an improvement in awareness of this behavior that involved limited cognitive effort, control, and intention [31]. This can be seen in the progressive increase of “No!” in the log and how the positive reinforcement also increased with the passage of time when stopping self-touches with the “Bravo!” strategy. These game dynamics respond to the need for control over self-touches, which must be trained through habits to block, divert, or perceive the act through attention and anticipation, as well as from the reflexive capacity of movement [2]. Although we perceive the touch when it has already occurred, with repetition, there is a possibility of anticipating and paying the necessary attention to stop the action before it happens.

Although the effectiveness of strategies using mindfulness [32] or using technological means, such as software [33], camera and sensor networks [34], and smartwatches [35], to reduce facial self-touches is being investigated, specific strategies should be provided to reduce self-touches, ranging from awareness to easily understanding triggers and mechanisms to regulate one’s own behavior, if possible, from everyday contexts [8,9,15,36]. This was the case in this intervention. One participant said as follows:

> The most important thing in this experience is to consider each situation that causes me to touch my face or the situations that do not. I have questioned every detail (makeup, fringe, daily face washing, piercings, glasses, hot-cold, allergies, etc.) [participant 14]

Thus, awareness is understood as a mental activity of the individuals that enables them to feel, have a reflexive knowledge of their actions, and know about themselves in particular, including actions less regulated by consciousness.

This serious game was considered useful, satisfactory, and fun—an intrinsic aspect of the act of playing [37]. This shows the positive balance of game design by providing a state of well-being, pleasure, and enjoyment, as described in flow theory [38], reinforcing the notion that serious games can be a playful and effective methodological framework for learning [39,40].

In a game aimed at learning, the media and materials are fundamental and enable its development [41], and if they are technological, they can be relevant allies for interaction [42], as the highly valued video tutorials, infographics, daily logs, and the Telegram group proved to be in this study. In summary, a game like #RedPingüiNO is based on its design and implementation, and therein lies its potential, taking care of the aspects of the mechanics, dynamics, and aesthetics framework. This mechanics, dynamics, and aesthetics framework is a formal approach to understanding games, which attempts to bring together game design and development, criticism, and technical research [43].

Starting from these elements, the intrinsic motivation of the participants was sought from Marczewski theory [44], valuing autonomy in game development, the mastery that enabled the participants to be aware of the progress they were achieving, the purpose for which they were playing, and their relationships. According to the latter, the participants from the groups both with and without reinforcement managed to be more aware of the times they touched their faces and perceived a reduction in self-touches. However, the participants in the EG-R considered the game more appealing, believed more firmly that the allotted time was sufficient, and presented higher frequencies of self-touch awareness and restraint during the recording. In summary, the game without this accompaniment was helpful,
although the collective support and reinforcements promoted among those who played it may have a relevant supporting role. Regarding playing with other people, a significant number of participants indicated that they had shared it with their families and friends, seeking out that social reinforcement themselves but within their own environment:

[What has struck me the most in this intervention has been the fact that I shared it with my family. They have helped me to be much more aware of the number of times I touch my face. [participant 8]

I have learned a lot, in fact, I have encouraged my family and friends to do it with me and together we have greatly reduced how much we touch our faces. When we sat down to eat, we talked about it and we kept a family ranking too. [participant 17]

Limitations and Future Directions

#RedPingüiNO resulted in a significant and helpful serious game that increased the perception of face self-touch avoidance. However, this study had some limitations. First, the sample used was a small number of university students, and it would be expedient to conduct future research with larger and more diverse samples. Second, owing to the specific characteristics of the university groups of students, there was a certain difference between the number of male and female participants in the different groups, and the CG had a smaller number of participants, which could bias the results. However, no significant gender differences were found in the study. Third, data on whether participants touched their face were not obtained from recordings. As not touching the face is a self-controlled behavior, these data were collected via a questionnaire and the daily logs of participants, and the results could be overestimated or underestimated. Therefore, as this behavior was not directly observed, these results must be understood as participants' perceptions of the decrease in the frequency of facial self-touches.

However, these limitations may present an opportunity because self-control itself has been one more strategy for reducing facial self-touches, with people taking a more active role in being aware of it and playing the game in their daily environments, not under laboratory conditions or in simulations. The perception of automated behaviors serves as a mechanism for self-control, which is facilitated by the awareness generated through conscious observation. Moreover, the perceived efficacy of the game invites to demonstrate its potential among other populations, such as health care workers and susceptible people, or within other environments.

Conclusions

This study provides evidence on the possible impact of the serious game #RedPingüiNO in reducing the behavior and perception of touching the face and its mucous membranes and may have an effect on the prevention of infectious diseases in different health contexts, such as those caused by SARS-CoV-2 and other pathogens. Being able to exert some control over this transmission path from within a game has allowed for better conscious control beyond verbally warning people not to touch their faces. The implemented tools were considered valid and reliable. Participants perceived a reduction in facial self-touches after the intervention and became more aware of their risk behaviors. The play strategy of #RedPingüiNO is simple and motivating in its format, precisely because it deals with a complex behavior that has a high degree of unconscious movement and automatic action, which is regulated by various contextual internal and external communication situations for the individual. Challenge-oriented mechanics, levels, and humor-based onomatopoeia reinforcement, as well as accessible and free gameplay components, make this serious game adaptable to other contexts such as health institutions, hospitals, community centers, schools, or similar entities. We also noted that the game could have a greater effect if it is shared, and the experience is reinforced by the interactions between the participants. #RedPingüiNO can help control risk behaviors against infectious diseases or in exceptional health situations. In summary, this study can be presented as a precedent for the self-control of behaviors related to health at different levels.

Acknowledgments

The authors wish to thank the Universidad de Alcalá for its collaboration and action during the COVID-19 pandemic. This project obtained a recognition diploma from the Universidad de Alcalá, a distinction for its contribution to alleviating the situation caused by the COVID-19 pandemic. The authors also wish to thank all the people and experts who collaborated in the development and dissemination of the project. This work was supported by the research project COVID-19 UAH 2019/00003/016/001/019, funded by the Universidad de Alcalá and Santander Universidades.

Authors’ Contributions

MAB designed and directed the study. MAB, PSO, and AVS performed the bibliographic review. MAB, PSO, AVS, and IAI structured the theoretical bases and developed the project. AVS organized the development of the instruments. MAB, PSO, and AVS performed the interventions. CJL developed the statistical analysis and structured the results. MAB, PSO, AVS, CJL, and IAI wrote the manuscript. MAB, AVS, and IAI reviewed the draft manuscript. MAB, PSO, AVS, CJL, and IAI approved the latest version of the manuscript.

Conflicts of Interest

None declared.
References


34. Senthilkumaran S, Arathisenthil SV, Meenakshisundaram R, Thirumalaikolundusubramanian P. Not touching the face is harder than it sounds: need for an intervention. Indian J Crit Care Med 2020 Aug;24(8):662-663 [FREE Full text] [doi: 10.5005/jp-journals-10071-23527] [Medline: 33024371]

Abbreviations

CG: control group
EG: experimental group
EG-nR: experimental group with no additional social reinforcement
EG-R: experimental group with additional social reinforcement
sFST: spontaneous facial self-touches
Eating Behavior and Satiety With Virtual Reality Meals Compared With Real Meals: Randomized Crossover Study

Alkyoni Glympi¹, MSc; Dorothy Odegi¹, MSc; Modjtaba Zandian¹, PhD; Per Södersten¹, PhD; Cecilia Bergh², PhD; Billy Langlet¹, PhD

¹Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Stockholm, Sweden
²Mandometer Clinics, Stockholm, Sweden

Corresponding Author:
Billy Langlet, PhD
Division of Clinical Geriatrics
Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels allé 23
Huddinge, Stockholm, 141 52
Sweden
Phone: 46 086905826
Email: billy.langlet@ki.se

Abstract

Background: Eating disorders and obesity are serious health problems with poor treatment outcomes and high relapse rates despite well-established treatments. Several studies have suggested that virtual reality technology could enhance the current treatment outcomes and could be used as an adjunctive tool in their treatment.

Objective: This study aims to investigate the differences between eating virtual and real-life meals and test the hypothesis that eating a virtual meal can reduce hunger among healthy women.

Methods: The study included 20 healthy women and used a randomized crossover design. The participants were asked to eat 1 introduction meal, 2 real meals, and 2 virtual meals, all containing real or virtual meatballs and potatoes. The real meals were eaten on a plate that had been placed on a scale that communicated with analytical software on a computer. The virtual meals were eaten in a room where participants were seated on a real chair in front of a real table and fitted with the virtual reality equipment. The eating behavior for both the real and virtual meals was filmed. Hunger was measured before and after the meals using questionnaires.

Results: There was a significant difference in hunger from baseline to after the real meal (mean difference=61.8, \( P < .001 \)) but no significant change in hunger from before to after the virtual meal (mean difference=6.9, \( P = .10 \)). There was no significant difference in food intake between the virtual and real meals (mean difference=36.8, \( P = .07 \)). Meal duration was significantly shorter in the virtual meal (mean difference=–5.4, \( P < .001 \)), which led to a higher eating rate (mean difference=82.9, \( P < .001 \)). Some participants took bites and chewed during the virtual meal, but the number of bites and chews was lower than in the real meal. The meal duration was reduced from the first virtual meal to the second virtual meal, but no significant difference was observed between the 2 real meals.

Conclusions: Eating a virtual meal does not appear to significantly reduce hunger in healthy individuals. Also, this methodology does not significantly result in eating behaviors identical to real-life conditions but does evoke chewing and bite behavior in certain individuals.

Trial Registration: ClinicalTrials.gov NCT05734209, https://clinicaltrials.gov/ct2/show/NCT05734209

(JMIR Serious Games 2023;11:e44348) doi:10.2196/44348

KEYWORDS

exposure therapy; eating behavior; anorexia nervosa; bulimia nervosa; binge eating disorder; overweight; obesity; immersive virtual reality; VR; virtual reality
Introduction

Eating disorders and obesity constitute severe health problems in Western society. Globally, the prevalence of eating disorders has increased recently by 25% [1], and for obesity, it is estimated that, by 2030, 20% of the world’s adult population will be obese [2]. Both health conditions considerably impair physical health, disrupt psychosocial functioning [1,3,4], and are risk factors for depression and anxiety disorders [3,4]. Both are expensive disorders, with the overall cost of eating disorders in the United States in 2018 estimated at $64.7 billion per year [5] and of high BMI worldwide in 2017 estimated to be $802 billion [6].

Although both health problems are related to food consumption, their commonalities are seldom discussed. For instance, the same eating behavior parameters (ie, eating rate and food intake) have been identified to increase the risk of developing anorexia and obesity [7,8]. A majority of these individuals have lost the ability to experience satiety during and after meals [7]. A key characteristic of anorexia nervosa is a slow eating rate, which stems from an effort to prevent weight gain by restricting food intake [7]. People with bulimia nervosa try to restrict their food intake to control their body weight but eventually become sufficiently hungry that they consume large amounts of food rapidly before purging [9]. Binge eaters have the bulimia intake pattern without purging [9]. Those who become obese often try to restrict their food intake, but their efforts are typically transitory and ineffective. Finally, a high eating rate is associated with increased food intake [8], excess body weight [8,10], and a larger bite size [8,11].

Eating disorders and obesity constitute severe health problems, and although their prevalence has increased over the years, there is still no consensus for their treatment [1]. The effectiveness of conventional eating disorder treatments, such as family-based therapy, psychopharmacological treatment, and cognitive behavior therapy, are weak [7], with most studies reporting poor long-term treatment outcomes [12]. Conventional treatment for obesity includes lifestyle modification (diet and exercise), medication, and bariatric surgery. The weight loss associated with lifestyle modification and medication interventions is also limited, due to poor adherence rates [13], and although bariatric surgery is the most effective weight loss treatment, it is associated with multiple health complications [14].

High relapse rates are reported for both eating disorders and obesity, with only some people with anorexia nervosa reaching full recovery [15] and people with obesity often regaining most of the lost weight [16]. Last, high dropout rates are another issue that affects the treatment outcomes. For instance, it is estimated that, for people with anorexia nervosa, the dropout rates range from 20.2% to 49.6% [17] and, for people with obesity, from 10% to more than 80%, depending on the treatment program [18]. Therefore, there is an urgent need to explore effective alternative treatments to address these health problems.

Virtual reality devices have been used in many areas of health care for a variety of applications from professional health care education to disease treatment [19]. Examples of conditions that have been treated successfully with virtual reality are anxiety disorders, phobias, stress management, posttraumatic stress disorder, acute and chronic pain management, autism, and physical rehabilitation [20,21]. An especially promising area of use may be for the assessment and treatment of eating disorders and obesity, for which virtual reality devices can be used as an adjunctive treatment tool [20,22,23].

Recent technological advancements have significantly decreased the cost of virtual reality, making the technology more accessible for medical treatments [24]. The ability to customize and control the virtual reality environment makes the application of this technology easier in a treatment context, since triggering certain fears may be difficult to reproduce in real life [25]. The scripted nature of a virtual reality app prevents the occurrence of unexpected situations and provides patients with a safe environment where they can experience specific aspects of life, allowing them to be “emotionally present” within the virtual environment [26]. Immersive virtual reality provides an even more intense experience for the user, using a virtual reality headset or multiprojected environments to generate realistic images, sounds, and other sensations [25].

Virtual reality exposure therapy has been successful in modifying body image distortions in people with obesity, binge eating disorder, and eating disorders not otherwise specified, leading to a reduction in inappropriate behaviors related to food and social relationships [27].

Besides modifying body image distortions, virtual reality studies of eating disorders have focused on exposure to virtual food stimuli [23]. Some of these studies aimed to assess emotional reactions or craving while using virtual reality, while others had therapeutic objectives [23]. The results suggest that the participants perceived the virtual reality environment as real and that it was able to elicit emotional responses (eg, anxiety, food-related thoughts) [23]. More specifically, for binge eating disorder and bulimia nervosa, studies indicate the potential utility of virtual reality in improving motivation for change, self-esteem, and reducing binge eating and purging behaviors [20].

In obesity, besides modifying body image distortion, virtual reality studies were successful at improving self-control, mood, and weight loss [28-30]. Another area of focus was to increase physical activity using “exergaming” as a potential treatment for obesity and diabetes. Exergaming’s fundamental concept is to use vigorous body activity as the input for interacting with digital game content, in the hopes of replacing the sedentary activity that is typical of traditional keyboard, game pad, and joystick inputs [31,32].

In addition to these findings, a study that investigated the perceived usability of a virtual reality app that promoted normal eating behavior suggested that clinicians have a positive attitude toward using virtual reality when treating people with an eating disorder [33]. Considering the advantages provided by this technology, virtual reality could be a useful tool to address eating disorders and obesity via exposure therapy and behavioral training.

Although interest in using virtual reality for treating eating disorders and obesity has increased over the last 2 decades, there is a lack of studies examining the effect of behavioral training
via virtual reality on eating behaviors and satiety [22,23]. It has been proven that better treatment outcomes are obtained when the treatment normalizes eating behavior [34]. Similar methods may be successful in treating other eating behavior problems. Thus, a behavioral training approach via virtual reality could provide a new adjunctive tool to the existing treatment of eating disorders and obesity.

Both the method (eg, virtual reality and immersive virtual reality) [23] and aim (eg, normalizing body image and food exposure) [22,23] of virtual reality studies in eating disorders and obesity have been inconclusive, and further work is required to clarify these findings. Another reason to investigate the effectiveness of virtual reality further is that clinicians have expressed concerns that people with anorexia nervosa could substitute virtual meals for real meals to reduce feelings of hunger without ingesting any food [33].

Therefore, the main aim of the study was to determine whether eating in an immersive virtual reality environment reduces hunger among healthy adults. The secondary aim was to describe any eating behavior differences between the consumption of virtual reality meals and real-life meals.

**Methods**

**Participants**

General inclusion criteria were women, 18 years to 28 years of age, a BMI from 18.5 kg/m$^2$ to 29 kg/m$^2$, not pregnant nor breastfeeding, and normal physical activity (<1500 metabolic equivalent of task min/week of moderate-to-vigorous physical activity level from the International Physical Activity Questionnaire). Health inclusion criteria were nonsmokers, the absence of temporomandibular disorders or recent serious dental surgery, not undergoing treatments known to affect appetite (eg, some psychotropic drugs, acetazolamide), and no previous history of eating disorders. Only nonvegetarians with no aversion to the food that was to be served were allowed to participate in the study.

**Recruitment**

Recruitment of participants started in October 2020 and ended in April 2022. Participants were recruited via digital meeting platforms, by contacting potential participants in person, and by hanging flyers around the university campus.

**Instruments**

**BMI**

To measure BMI, weight was measured using a body composition analyzer (Tanita BC-418), and height was measured using a stadiometer (Seca 216 1814009). To collect data on eating behavior (ie, chews, bites, forkfuls) a home security camera (Mi 360°, Xiaomi) was used. To display and interact with the virtual meal, we used a virtual reality headset (HTC VIVE Pro), which consists of a head-mounted device and accompanying handheld controllers. The headset provides 110 degrees of viewing at a resolution of 1440 x 1600 pixels per eye (2880 x 1600 pixels combined) and a refresh rate of 90 Hz.

**Food Intake**

To measure food intake throughout the real meal, a digital food scale (Mandometer version 4) was used. The Mandometer is a medical device that consists of a scale that is connected to a computer; the scale measures the weight of the food on the plate at a sampling frequency of 1 Hz [35]. Before starting a meal, a plate was placed on the food scale, and the weight of the plate was subtracted from the meal. Food was then served on the plate, and the weight of the food was registered throughout the meal.

**Subjective Data**

Subjective data were collected via a tablet (Samsung Galaxy Tab A7) using an in-house developed app containing all questionnaires to ensure that participants fulfilled inclusion criteria, and it also contained 1 questionnaire to collect outcome data.

**Inclusion Questionnaires**

The short form of the International Physical Activity Questionnaire was used to evaluate participants' physical activity and to exclude those with too high moderate-to-vigorous physical activity levels [36]. The questionnaire consists of 7 questions, for example “During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?”. The Dutch Eating Behavior Questionnaire was used to assess 3 distinct eating behaviors (external, emotional, and restrained eating) [37]. The questionnaire consists of 33 questions, such as “If you have put on weight, do you eat less than you usually do?”, “Do you have the desire to eat, when you are emotionally upset?”, or “If you see others eating, do you also have the desire to eat?”

**Outcome Questionnaire**

Subjective appetite, mood, and taste of the food were rated using a meal questionnaire, which consists of 25 questions divided into 2 parts. The first part, “before meal,” consists of 6 questions about mood (eg, “Are you feeling tense right now?”) and 6 questions about appetite (eg, “How hungry do you feel?”). Mood and appetite questions were rated on 100-mm visual analog scales ranging from 0 to 100, with the verbal anchors “Not at all”=0 on the left side and “Extremely”=100 on the right side. The second part, “after meal,” consists of 6 questions about appetite (same as before the meal) and 7 questions about the taste of the food (eg, “How fatty did the food taste?”) on visual analog scales identical to that presented before the meal.

**Ethical Approval**

The research protocol was approved by the Swedish Ethical Review Authority (D.nr: 2019-04249).

**Procedure**

The study consisted of 5 meetings in a randomized crossover design, in which participants acted as their own control with 2
conditions, each with 2 repeats. Respondents first attended an information meeting where they ate an introduction meal. During the remaining 4 meetings, participants were either served a virtual or a real meal in a randomized order (Figure 1). All 5 meetings, including the information meeting, were held on weekdays (Monday to Friday) during lunch hours (11 AM to 1 PM) and had a wash-out period of at least 3 days. Both the real-life and virtual reality meals were recorded using a video camera.

Figure 1. Study design.

The day before each meeting, participants received an email with information regarding the type of meal (virtual or real) that they would have. To control for satiety, participants were reminded to eat the same breakfast the day of the meeting, at least 3 hours before, and to refrain from high-intensity physical activity 24 hours before the meeting. All the real meals were prepared by the researchers in the clinic 30 minutes to 60 minutes before the participants’ arrival. All the prepared meals were kept in the oven at a temperature of 80 °C to ensure proper serving temperature. In the virtual world, the size of utensils, food containers, and food items were equivalent to the matching real-world item. Size-to-weight of food items in the virtual world was calculated, and corresponding meatballs and potatoes were purchased for the real-world meals, resulting in meatballs weighing 14 g each and potatoes weighing 60 g each.

Information Meeting

During the information meeting, respondents were provided with detailed information about the study, data collection, handling and secrecy, and the potential risks associated with participating. Besides informing them, the purpose was to ensure their eligibility to participate in the study, familiarize them with the study procedures, and provide a baseline food intake value for the rest of the meals. Each participant signed a written consent form and proceeded to answer the Health and the International Physical Activity Questionnaires. Once the questionnaires were reviewed and approved, the participant’s height and body composition were measured.

Afterward, the introduction meal was served. In total, 1.2 kg of food was served, consisting of meatballs (400 g) and boiled potatoes (800 g). Brown sauce and lingonberry jam were placed in separate ceramic bowls on the table with a glass of water (200 mL). After the food had been served, the participant filled in the “before meal” part of the meal questionnaire. Once completed, the participant was instructed to put food on the plate that had been placed on the food scale and to eat the meal as they normally would, with no time nor intake restriction (ad libitum). Finally, instructions were given for the participant that they could add food to the plate or leave food on the plate and to refrain from using a mobile phone or other distracting elements while eating. Once the participant was finished putting food on the plate, the food scale and the video camera were turned on. After meal termination, the participant filled in the “after meal” part of the meal questionnaire.

The next step of the procedure was the virtual meal, for which the participant was taken to the virtual reality room, seated on a real chair in front of a real table, and fitted with the virtual reality equipment. The virtual reality room showed a kitchen that included a table, chair, and windows that revealed a natural scene. In the virtual reality environment, the controllers were represented by hands. The introduction meal consisted of several steps, teaching the participant how to serve themselves and consume food in the virtual reality environment [33]. A virtual tablet presented what tasks to perform to proceed to the next step. In every step, a new “object” appeared on the table starting with a plate and continuing with meatballs in a pan, potatoes in a pot, brown sauce in a sauceboat, lingonberry jam, a jug of water, a glass, and cutlery [33]. After completing all previous steps, the participant could freely “cut” and “eat” the food.

Real-life Meal

The meeting for the real-life meals started by measuring the participant’s body composition and filling in the “before meal” part of the meal questionnaire. The participant was then led to the room where they would eat. The room was designed to be as neutral as possible, with only a table and a chair (Figure 2). The meal was served on a plate that was placed on the food scale, with a glass of water alongside. The quantity of food on the plate was the same as the amount of food that the participant consumed during the introduction meal, plus an extra 50% portion on the side (Figure 3). The same instructions as in the introduction meal were provided. Once the participant was done
eating, they filled in the “after meal” part of the meal questionnaire.

**Figure 2.** Research room.

![Research room image](image1)

**Figure 3.** Food was served, and 50% of the portion was placed on the side.

![Plate with food](image2)

**Virtual Meal**

The meeting started by measuring the participant’s body composition. Afterward, the participant was taken to the virtual reality room (**Figure 4**), filled in the “before meal” part of the meal questionnaire, and was fitted with the virtual reality equipment. The meal served in virtual reality was the “same” as the real meal; the portion of the virtual meal was equivalent to the portion that was consumed during the introduction meal.

There was extra food in the containers placed on the table, enabling free addition of food to the plate (**Figure 5**). Once the participant was ready to eat the virtual meal, the video camera and the acquisition software from the computer were initiated by the researcher. The purpose was for the meal to be eaten normally with no time restriction. Once the participant was done, the “after meal” part of the meal questionnaire was filled in.
Data Formatting

Video recordings from both virtual and real meals were annotated using The Observer XT 16 (Noldus). The eating-related behaviors annotated were forkful, food addition, bite, and chew. Forkful was annotated as a point event when the fork touched the plate to pick up food. Addition was annotated as a state event, starting when food was added to the plate and ending when no more food was added and food ingestion started again. Bite was annotated as a point event when the mouth was closed around the fork. The virtual food did not require bites, as the food was “eaten” by putting it close to the headset. Chew was annotated as a point event when teeth occlusion and activation of masseter and temporalis muscles occurred.

Statistical Analysis

Statistical analyses were carried out using R software version 4.2.1 [38]. Since participants conducted 2 meals for each eating condition (2 real and 2 virtual meals), the average of every measured outcome (eg, food intake, hunger) was calculated. Thus, for each eating condition, 1 value (the mean) was created and was used in the analysis. For example, the food intake of the virtual reality meal refers to the mean food intake from the first and the second virtual reality meals.

The Shapiro-Wilk test and visual inspection of a Q-Q plot were performed to ensure normal distribution of the data. For the parameters that were not normally distributed, a Wilcoxon signed rank test and a paired Student $t$ test were performed. Since the interpretation (ie, rejection or nonrejection of the null hypothesis) to both tests were identical for all investigated
parameters, only *t* tests are presented in the results. The relevant parameters tested were hunger before and after the virtual and real meals, food intake, meal duration, eating rate, forkfuls, and additions, between the real and virtual meals. A subgroup was created from the participants who took chews and bites in the virtual meals. Additional *t* tests were performed for this subgroup, comparing chews and bites in the virtual meal to the real meal. Last, the same test was performed for all participants to evaluate the effect on hunger and meal duration between the first and second meals in both eating conditions. Pearson product-moment correlation was used to examine the association between hunger after the meal and meal duration and chews, as well as the correlation between meal duration and chews and food intake. Caution is required when interpreting all correlation analyses including chews from the virtual reality condition as a parameter, since in more than one-half of the measurements, the number of chews was 0. A significance threshold of $P \leq .05$ for all statistical tests was used. Unless otherwise specified, values are expressed as mean (SD). For the power calculation, a sample size of 20 was found to be sufficient for the study, using the effect size between real meal conditions of previous studies, an $\alpha$ cutoff of .05, and power of 0.80.

## Results

### Participants

Of 24 participants, 20 finished all parts of the study (Table 1); 2 participants dropped out due to scheduling difficulties, 1 participant was excluded due to corrupt data, and 1 participant was excluded due to suspected conformity bias, based on low food intake. No participant was excluded based on questionnaire responses.

### Hunger Before and After Meals

For the virtual meal, participants rated their hunger before and after the meal at 66.9 and 59.9 (out of 100), respectively, resulting in no significant difference in hunger before and after the virtual meal ($P = .10$). For the real meal, participants rated their hunger before and after the meal at 68.6 and 6.9, respectively, resulting in a significant reduction in hunger after the real meal (mean difference = 61.8, $P < .001$). One participant had a reduction in hunger similar in size to what was expected for a real meal as a result of eating a virtual meal, with a rated reduction of 59.

There was no significant difference in hunger before the meal between the virtual and real meals ($P = .79$), but hunger was significantly higher in the virtual condition after the meal (mean difference = $-53.1, P < .001$).

### Eating Behavior Characteristics

There was no significant difference in “food” intake between the virtual and real meals ($P = .07$). Meal duration was significantly shorter with the virtual meal (mean difference = $-5.4, P < .001$), which led to a higher eating rate (mean difference = $82.9, P < .001$) for the virtual meal. When the chewing time (burst interval) was removed from the virtual and real meals, there was no significant difference in meal duration between the virtual and real meals ($P = .06$). The number of forkfuls and number of additions were significantly lower with the virtual meal than with the real meal (mean difference = $-13.0, P < .001$ and mean difference = $-2.0, P = .005$, respectively). The results are presented in Table 2.

### Table 1. Demographic data of the participants (n=20).

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Results, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>23.9 (2.5)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>61.2 (9.6)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>165.3 (7.4)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>22.4 (2.8)</td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>27.5 (5.1)</td>
</tr>
</tbody>
</table>

### Table 2. Eating behavior characteristics from the virtual and real meals (n=20).

<table>
<thead>
<tr>
<th>Eating behavior</th>
<th>Virtual meal, mean (SD)</th>
<th>Real meal, mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food intake (g)</td>
<td>375.6 (140.9)</td>
<td>338.9 (108.3)</td>
<td>.07</td>
</tr>
<tr>
<td>Meal duration (min)</td>
<td>3.6 (1.7)</td>
<td>9.0 (3.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Meal duration without chews (min)</td>
<td>2.8 (1.3)</td>
<td>2.4 (1.2)</td>
<td>.06</td>
</tr>
<tr>
<td>Eating rate (g/min)</td>
<td>124.0 (52.9)</td>
<td>41.1 (16.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Forkfuls (n)</td>
<td>25.9 (9.8)</td>
<td>38.9 (14.8)</td>
<td>.002</td>
</tr>
<tr>
<td>Additions (n)</td>
<td>0.6 (0.8)</td>
<td>2.5 (2.9)</td>
<td>.005</td>
</tr>
<tr>
<td>Bites (n)</td>
<td>10.0 (11.7)</td>
<td>40.0 (15.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chews (n)</td>
<td>28.9 (68.0)</td>
<td>419.0 (162.9)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Chews and Bites
For participants who took bites and chews in the virtual meal, the numbers were significantly lower compared with the real meal (n=14, mean difference=-28.1, P<.001 and n=6, mean difference=-387.4, P=.004, respectively). The results are presented in Table 3.

<table>
<thead>
<tr>
<th>Eating behavior</th>
<th>Responding (n=20), n (%)</th>
<th>Virtual meal, mean (SD)</th>
<th>Real meal, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bites</td>
<td>14 (70)</td>
<td>15.4 (11.2)</td>
<td>42.4 (16.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chews</td>
<td>6 (30)</td>
<td>82.4 (97.5)</td>
<td>469.8 (193.0)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Number included in the analysis, based on the participants who took bites or chews in the virtual meal.

Familiarization Effects
There was a significant reduction in meal duration from the first virtual meal to the second virtual meal (mean difference=0.4, P=.03) but no significant change in meal duration between the first real meal and second real meal (P=.07). There was no significant difference in hunger before the meal between the first virtual meal and second virtual meal and between the first real meal and second real meal (P=.40 and P=.22, respectively).

Association Between Hunger and Chews and Meal Duration
In both the virtual and real meals, a low correlation was found between hunger after the meal and number of chews (r=-0.10, P=.68 and r=-0.18, P=.44, respectively). A low correlation was also found between hunger after the meal and meal duration in both the virtual and real meals (r=-0.12, P=.60 and r=-.12, P=.62, respectively).

Association Between Meal Duration and Chews and Food Intake
In both the virtual and real meals, a low correlation was found between meal duration and food intake (r=0.22, P=.35 and r=0.14, P=.56, respectively). A high correlation was found between meal duration and chews in the virtual and real meals (r=0.69, P<.001 and r=0.75, P<.001, respectively).

Discussion
Principal Findings
To our knowledge, this is the first study to investigate the effect of a virtual meal on hunger and to explore the differences in eating behavior between real and virtual meals. The findings of this study indicate that eating a virtual meal does not affect hunger in healthy women. The results demonstrated that participants consumed approximately the same amount of real and virtual “food,” but the virtual meal was significantly shorter, resulting in a higher eating rate. Participants also had lower numbers of chews, bites, forkfuls, and additions in the virtual meal.

Comparison With Prior Work
It is expected for hunger to decline throughout a real meal, which was the case for the real meal condition. Meanwhile, eating a virtual meal did not significantly reduce hunger. The nonsignificant difference in hunger before and after a virtual meal suggests that such meals can be used for people with anorexia nervosa without the risk of having them replacing real meals with virtual meals. This finding is in contrast with a study in which participants were found to decrease their consumption of a particular food (eg, cheese) by repeatedly imagining eating it [39]. Since this study was conducted with healthy women, care will be required when implementing this method for eating disorders or obesity.

The extent to which a virtual meal represents realistic eating conditions was also examined, although the aim of the virtual reality app was not to mimic real-life eating behavior. Interestingly, even though the meal duration of the virtual meal was shorter than that of the real meal, the amount of “food” consumed was approximately the same. The reason for this phenomenon was that most of the participants did not chew during the virtual meal, resulting in an approximately 3 times higher eating rate compared with the real meal. These data support previous research, which showed that eating rate was inversely related to the number of chews [40]. Changing the virtual reality protocol to elicit a more realistic eating behavior or providing virtual reality eating training may result in a satiety response. However, findings suggest that this approach would not be beneficial as an intervention for obesity or bulimia nervosa.

A serendipitous finding of the study was that, although it was not required nor requested, participants took bites and chews during the virtual meals, indicating that the virtual reality environment was perceived on some level as being real. The display of eating behaviors similar to real life is in line with the findings that cue exposure in virtual environments is effective for inducing food craving in healthy control participants [41].

The familiarization effect observed in the virtual meals suggests more meals are required to reach a stable behavior since several participants experienced difficulties. Examples of those difficulties were placing the food in their mouth and grabbing or serving the food on their plate. People are less accustomed to eating virtual meals than real meals, and that conclusion is also supported by a study that examined the impact of the environment on eating behavior using virtual reality [42].

Based on previous findings, we expected longer meals (ie, longer exposure to food “images”) [43] and more chews (ie, more oral processing) [44] to be negatively associated with hunger. However, the correlation was low in both cases. These correlations should be interpreted with care, since they were not part of the initial hypothesis (we did not know participants would chew virtual food) and the fact that more than one-half of the participants did not chew in the virtual meal.
Strengths and Limitations
The main strength of this study was the use of immersive virtual reality equipment, as this technology increases the feeling of being present in the environment [20]. Other strengths were that many confounding covariates were controlled through the deployment of a random crossover design [45], narrow inclusion criteria, a strict study protocol [42], a familiarization meal, and objective data collection methods. For the introduction meal, a larger portion (1.2 kg of food) than what is typically ingested was offered to the participants to make them realize they did not have to consume everything. For the coming meals, they were provided with the same portion size as they ate during the introduction meal with extra food on the side. The artificial manipulation of portion size ensured similar meal sizes, making hunger, eating rate, and other eating behavior parameters comparable [46].

The main limitation of the study was the novelty of immersive virtual reality. Participants unused to immersive virtual reality require practice and concentration to understand the “game” mechanics [42]. It was observed that, in the introduction meal, several participants experienced difficulties handling the food (eg, placing the food in their mouth and grabbing or serving the food on their plate). Some participants expressed surprise or amusement during their encounter with virtual reality equipment (especially the first time), which may have influenced the results. Another limitation was that the virtual reality app required food on the plate and in containers to start, which resulted in alterations to the real meal protocol as well. Ideally, participants would have only received food on the plate for each session.

Other limitations include a lack of effect size measurements from previous virtual reality meals for the power analysis and potential low generalizability to people with eating disorders or obesity, because the study enrolled only healthy women. Moreover, the recruitment location (university campus) may have resulted in a convenience sample, in regard to education, socioeconomic status, technological literacy, and availability (workdays from 11 AM to 1 PM). In addition, the subjective methods used to quantify hunger are vulnerable to demand bias, in which the participants provide responses that they believe the researchers are looking for [43]. However, there are no reliable biological markers to evaluate appetite [47].

Future Perspectives
Further investigation is necessary to confirm the safety of virtual reality for patients, as they may respond differently than healthy control participants. Future studies should aim to include other demographic groups (eg, men, children, vegans, and other groups prone to eating disorders or obesity) to evaluate the generalizability of the findings. The effect of a virtual meal should also be investigated in different settings (eg, at home, at a restaurant, lunch with friends) and with a variety of foods. Moreover, future studies should provide additional virtual meals to reduce the familiarization effect observed in this study. Looking to the future, further attempts at incorporating chewing in the virtual meal (eg, with chewing gum to mimic food mastication) and using more advanced equipment that allows the involvement of all the senses (eg, sounds and olfactory information) might prove beneficial in evoking physiological responses similar to those of real meals.

Conclusions
This study investigated the effect of virtual reality on hunger. Eating in an immersive virtual reality environment has no significant effects on hunger in healthy individuals, suggesting that virtual reality can be used for patients (eg, with anorexia nervosa) without the risk of having them replace real meals with virtual meals. Due to the lack of chewing, the virtual meal was consumed in less time, while the food intake was approximately the same as in the real meal. Surprisingly, some participants took bites and chews in the virtual meal, although they were not instructed to nor was it required to “eat” the food.

Acknowledgments
The material and methods used and the collection of data used in this study were financially supported by the Mandometer clinic. The funding organization had no role in designing the study, collecting, and analyzing the data, or in preparing this manuscript.

Conflicts of Interest
PS and CB own shares in Mandometer AB, the company that financed the development of the virtual reality app. Other authors have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 359 KB - games_v11i1e44348_app1.pdf]

References
Developing a Game (Inner Dragon) Within a Leading Smartphone App for Smoking Cessation: Design and Feasibility Evaluation Study

Justin S White¹,²,³, PhD; Marie K Salem¹, MPH; Séverine Toussaert⁴, PhD; J Lee Westmaas⁵, PhD; Bethany R Raiff⁶, BCBA-D, PhD; David Crane⁷, PhD; Edward Warrender⁷; Courtney Lyles⁸, PhD; Lorien Abroms⁹, ScD; Johannes Thrul¹⁰,¹¹,¹², PhD

¹Philip R Lee Institute for Health Policy Studies, University of California, San Francisco, CA, United States
²Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, United States
³Department of Health Law, Policy & Management, Boston University School of Public Health, Boston, MA, United States
⁴Department of Economics, University of Oxford, Oxford, United Kingdom
⁵Population Science, American Cancer Society, Atlanta, GA, United States
⁶Department of Psychology, Rowan University, Glassboro, NJ, United States
⁷23 Limited, London, United Kingdom
⁸Department of Medicine, University of California, San Francisco, CA, United States
⁹Department of Prevention and Community Health, George Washington University, Washington, DC, United States
¹⁰Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
¹¹Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, United States
¹²Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia

Corresponding Author:
Justin S White, PhD
Philip R Lee Institute for Health Policy Studies
University of California
490 Illinois Street, Box 0936
San Francisco, CA, 94158
United States
Phone: 1 415 476 8045
Email: Justin.White@ucsf.edu

Abstract

Background: Several stand-alone smartphone apps have used serious games to provide an engaging approach to quitting smoking. So far, the uptake of these games has been modest, and the evidence base for their efficacy in promoting smoking cessation is still evolving. The feasibility of integrating a game into a popular smoking cessation app is unclear.

Objective: The aim of this paper was to describe the design and iterative development of the Inner Dragon game within Smoke Free, a smartphone app with proven efficacy, and the results of a single-arm feasibility trial as part of a broad program that seeks to assess the effectiveness of the gamified app for smoking cessation.

Methods: In phase 1, the study team undertook a multistep process to design and develop the game, including web-based focus group discussions with end users (n=15). In phase 2, a single-arm study of Smoke Free users who were trying to quit (n=30) was conducted to assess the feasibility and acceptability of the integrated game and to establish the feasibility of the planned procedures for a randomized pilot trial.

Results: Phase 1 led to the final design of Inner Dragon, informed by principles from psychology and behavioral economics and incorporating several game mechanics designed to increase user engagement and retention. Inner Dragon users maintain an evolving pet dragon that serves as a virtual avatar for the users’ progress in quitting. The phase-2 study established the feasibility of the study methods. The mean number of app sessions completed per user was 13.8 (SD 13.1; median 8; range 1-46), with a mean duration per session of 5.8 (median 1.1; range 0-81.1) minutes. Overall, three-fourths (18/24, 75%) of the participants entered the Inner Dragon game at least once and had a mean of 2.4 (SD 2.4) sessions of game use. The use of Inner Dragon was positively associated with the total number of app sessions (correlation 0.57). The mean satisfaction score of participants who
Background

Cigarette smoking is a leading preventable cause of death and disease in the United States [1]. Although many clinic-based and phone-based smoking cessation treatments are effective, most people who try to quit smoking do so without any aids and are successful <5% of the time [2]. The recommended cessation treatment, which combines counseling with medication, is used by <5% of those who attempt to quit [3]. Smartphone apps for smoking cessation have become increasingly popular among smokers who want to quit. As of 2020, English-language smoking cessation apps had been downloaded 33 million times [4]. The low cost and convenience of smartphone-delivered support makes them an appealing alternative to clinical treatments for many smokers, especially those disproportionately affected by the harms of smoking such as low-income, less educated, and racially minoritized subgroups [5]. In addition, the widespread availability of smartphones—85% of US adults owned one as of 2021—has helped extend the reach of smoking cessation apps to a large number of people [6]. Thus, smartphone-based interventions hold great promise for reducing the high burden and disparities in smoking-related cancer risk and mortality. Although many cessation apps do not follow evidence-based guidelines [7-9], a few are now being developed using established behavior change theory and existing clinical guidelines.

A perennial challenge for mobile health (mHealth) apps is their low engagement and retention rates [10]. Overall, <5% of users actively use a downloaded mHealth app daily, and <5% of users continue to use the app after 10 days [11]. Approximately half of the participants in studies of app-based mHealth interventions drop out, and studies of app-based smoking cessation interventions have found similarly high dropout rates [12-15]. Given the potential for smartphone apps to reach millions of smokers and reduce tobacco use, it is important to find ways to engage app users and increase retention and efficacy.

Behavioral researchers have suggested using serious games and gamification techniques to promote behavioral health [16-18]. Serious games are games with a primary purpose other than entertainment, such as health promotion. Gamification is a related motivation tool that uses nonmonetary rewards to make nongame activities fun or challenging. Gamification is designed to increase users’ motivation and engagement, which in turn may increase exposure and adherence to a smoking cessation program [19]. A study found that approximately 75% of the adult smokers surveyed played video games [20], and many smokers believed that game-based interventions for smoking cessation would be motivating [20,21].

Several gamified apps have been developed to help people quit smoking, such as Cigbreak [22], Crush the Crave [23,24], Inspired [25], QuitIT [26,27], Quittr [28,29], and Tobstop [30,31]. A small number of randomized controlled trials of these apps have been conducted, and they have generally found that users were satisfied with the gamified app, but most did not experience significant increases in smoking abstinence [23,27,29,31]. User engagement and retention of these apps also varied, with each study adopting a different approach to measurement (eg, number of game “episodes” completed or minutes played). Nonetheless, nonrandomized studies have suggested that elements of gamification, such as achievement badges, may increase users’ self-efficacy and motivation or intention to quit [32,33]. Taken together, these findings demonstrate limited and conflicting information regarding the effectiveness of gamified apps for increasing smoking cessation relative to quitting without assistance, and additional efforts are needed to improve user engagement and retention.

The Smoke Free App

Our team of academics and software developers sought to design and integrate a game into the Smoke Free smartphone app. Developed by London-based 23 Limited, Smoke Free is one of the most downloaded smoking cessation apps in the Apple and Android stores, with >800,000 downloads per year and 6 million downloads so far [34,35]. A preprint of a pragmatic randomized controlled trial reported that, although being randomized to receive an offer to use the Smoke Free app did not increase smoking abstinence compared with no intervention, the app increased continuous abstinence at 6 months when smokers who were randomized to receive the app downloaded it (12.7% intervention vs 7% comparator) [35]. The app leverages behavior change techniques that have proven to be effective in face-to-face behavioral support programs [36]. Key features of the app include (1) a calculator that tracks the total amount of money saved by not smoking; (2) a calendar that tracks the time elapsed since the user quit smoking; (3) a scoreboard that awards badges to users for not smoking; (4) progress indicators that inform users about health improvements that the user can expect because they started their quit attempt; (5) daily missions that assign evidence-based tasks to help users avoid and resist urges to smoke; (6) a diary and cravings log that track the frequency, strength, and location of cravings to smoke; and (7) a text-based chatbot that delivers quitting guidance in a friendly, conversational tone. In randomized trials, a full version of the
app with chatbot was shown to increase user engagement compared with a reduced version of the app [37], and the daily missions were shown to increase user retention and self-reported smoking abstinence at 3 months [34]. The current standard version of the Smoke Free app does not contain gamification features apart from the virtual badges noted previously.

The Inner Dragon Game

Our team designed a theory-informed game called Inner Dragon and integrated it into the Smoke Free smartphone app. Inner Dragon uses game elements that incorporate behavioral insights to increase user engagement and retention. This increased engagement is expected to increase exposure and adherence to the smoking cessation program, which should improve the chances of quitting smoking (Figure 1). The process of developing Inner Dragon and the game design is described in the Methods section.

This feasibility trial is a precursor to a 2-arm pilot randomized controlled trial and part of a broad program that seeks to assess the effectiveness of gamifying an app for increasing smoking cessation. The pilot trial will evaluate the engagement, satisfaction, and early efficacy of gamifying the Smoke Free app compared with a standard version of Smoke Free to guide the design of a large trial. The aim of this study was to assess the feasibility and acceptability of adding the Inner Dragon gamification component to the existing Smoke Free smartphone app for smoking cessation.

Methods

Overview

The study was conducted in 2 phases. In the first phase, the study team designed and integrated the game into the Smoke Free app. In the second phase, the study team evaluated the feasibility and acceptability of the integrated game using a single-arm trial. An overview of the study approach is provided in Figure 2. Our approach follows the biphasic framework for gamification design by Marczewski [38], involving a first phase of planning and design based on iterative information collection and development and a second phase of measuring user activities and engagement [39].

Phase 1—Game Design and Development

We followed a multistep usability testing procedure for the development of the Inner Dragon game: (1) we created a game design document and sketches of app screens for the game; (2) we conducted focus group discussions (FGDs) to gather feedback about sketches with Smoke Free users who are not participating in the feasibility trial; and (3) we finalized the game prototype, integrated it into Smoke Free, and performed internal app testing.

Game Development

As an initial step, we partnered with the game design studio, Workinman Interactive, LLC, to produce a game design document that described the significant features of the game. Workinman has a long history of developing games for apps and clients include Disney, Fisher-Price, and NBCUniversal. The investigative team provided an initial concept for the game. The game design document was refined based on a series of meetings and multiple rounds of revisions. Workinman also provided supporting art for wireframes, game flowcharts, and screen images. These materials were used by 23 Limited to build an initial prototype of the Inner Dragon game.

Design of Inner Design

In the Inner Dragon game, users care for a customizable pet dragon, whose growth reflects the user’s own progress in quitting smoking. The game uses traditional virtual pet retention mechanics with some social features and many options for customization and personalization. Virtual pets have been
Inner Dragon includes several game mechanics designed to increase user engagement and retention, with the primary goal of increasing exposure and adherence to the smoking cessation program, thereby improving the likelihood that the user quits smoking. Inner Dragon allows users to interact with a digital pet that is tied to their quit attempt as a means of progress toward continued abstinence, positive reinforcement for engagement with the smoking cessation program, and entertainment [40]. Game features were designed to accommodate any intensity of game play, whether heavy or sporadic, with no limitation on the amount of time or engagement. The key game mechanics were as follows. First, the user maintains a pet dragon that hatches on their quit day and evolves every 7 days to unlock new attributes and powers (Figure 3). The dragon acts as a virtual avatar that represents the user’s progress. Second, the user earns experience points by engaging in selected in-app activities, including those in the game (eg, playing a minigame or feeding the dragon) and in the core Smoke Free app (eg, completing a mission or logging a diary entry). Rewards for engaging with core app features were designed to improve adherence to the smoking cessation program. The experience points unlock gifts and cosmetics, directly rewarding frequent and consistent use of the game. The user can customize their dragon (eg, wing shape or clothing accessories) over the course of the quit experience by steadily unlocking features. Third, the Inner Dragon home screen has “care meters” that users must work to keep from falling very low: calmness, nutrition, hygiene, and energy. Engaging with the dragon in various ways increases the meters. For example, petting the dragon increases calmness, and feeding the dragon increases nutrition. Caring for and interacting with a virtual pet through the care meters can foster a bond with the pet and motivate users to return to the game regularly [40]. Fourth, the game provides tools for the user to better cope with the challenges of withdrawal: a dragon-led breathing exercise to provide calmness and a memory minigame as a distraction. Fifth, a user can asynchronously interact with other users’ dragons in a “dragon park” by (1) viewing their profile and progress and (2) sending and receiving motivational messages and emojis from a preset menu (Figure 4). These game mechanics are hypothesized to combine to increase the users’ engagement with the app and, subsequently, their chance of quitting successfully.

The game design was informed by principles from the fields of psychology and behavioral economics. As noted in the following section, aspects of the game were motivated by several theories and theoretical constructs, such as self-determination theory, theory of present bias and hyperbolic discounting, theory of salience and limited attention, and variable reinforcement under operant conditioning. The avatar provides salient, visual feedback with endogenous value (tied closely to the user’s motivation to quit) associated that may sustain and enhance motivation to quit. Self-determination theory predicts that this type of feedback is highly intrinsically motivating [42]. Furthermore, the user may identify with the digital pet as an avatar, and this may cultivate a digital therapeutic alliance with the game and app, for example, by creating a bond with the dragon and increasing the user’s confidence to succeed [43,44]. The use of frequent, salient, in-game rewards was designed to counter the behavioral economic constructs of present bias and inattention to app use. The design further included evidence-based practices from contingency management, such as the use of escalating in-game rewards for abstinence, with a reset point for lapses and sustained abstinence (by harnessing regret and loss aversion) [45]. The use of surprise gifts provided a variable reward structure designed to boost engagement and novelty. The asynchronous interactions with other users in the dragon park provide opportunities for limited social support and social comparisons that may motivate the user to exert more effort in the quit attempt [46,47].

The various game mechanics and elements were designed to appeal to users with different motivations. For example, Yee [48] identified 3 main components of player motivation: achievement (advancement and competition), social (socializing and relationship), and immersion (discovery, customization, and escapism). The experience point system may appeal to achievement seekers; the interaction with other players and a sense of connection with the dragon may appeal to socializers, and the rich opportunities for dragon customization and distraction game may appeal to players seeking immersion.

popular with consumers because they can foster bonding and companionship [40]. The Inner Dragon concept is also broadly compatible with the genres of games most liked by adult smokers, namely action, role-playing, and action-adventure games [41].
Figure 3. Dragon evolution. (A) Egg hatches into full dragon. (B) Evolves every 7d, gains new “abilities.”.

Figure 4. Dragon care. (A) Feeding dragon (fun). (B) Memory mini-game (distraction). (C) Breathing exercise (manage cravings). (D) Dragon park (social feedback and support).

FGD Procedures

Following the development of an initial game prototype, the investigative team conducted FGDs with existing users of the Smoke Free app to gather their feedback about the Inner Dragon game concept. The findings from the FGDs were used to refine Inner Dragon in preparation for the feasibility trial.

FGD participants were recruited using a recruitment message displayed to new Smoke Free users on an onboarding screen. The message included a hyperlink to a Qualtrics screening questionnaire. Eligibility criteria included the following: aged ≥18 years; had downloaded and opened the Smoke Free app; lived in the United States; spoke, wrote, and read English; smoked at least 1 cigarette every day; and was available during one of the scheduled FGDs.

The study team organized three 90-minute FGDs via Zoom (Zoom Video Communications). During each session, a facilitator guided the participants through several discussion topics: past smoking experience, preferences for methods of quitting, and feedback about the game prototype (Multimedia Appendix 1). For the latter topic, a facilitator provided the FGD participants with a description of the different elements of the initial prototype and accompanying screenshots. Participants received a payment of US $40 as a choice of a gift card or cash card provided through the Tremendous digital payment platform.
FGD sessions were transcribed. Overall, 2 independent coders identified key themes from the transcripts. Coders then input their findings into a series of “feedback capture grids,” which organized the likes, criticisms, questions, and ideas of users for each discussion topic. Each gamification feature was evaluated with respect to its interest, ease of use, motivational potential, and overall satisfaction.

Phase 2—Feasibility Testing

Overview
This study included a single-arm, unblinded evaluation of the feasibility and acceptability of the Smoke Free app with integrated Inner Dragon game, in preparation for a planned randomized pilot trial. All participants in the feasibility study were given free access to the Smoke Free app with the integrated Inner Dragon game. Participants’ use of the app was passively monitored for 28 days, and follow-up assessments occurred after 1 month.

Recruitment and Participants
Study participants were recruited from the population of general users of the Smoke Free app. Focus group participants were not included. New users of Smoke Free were invited to participate in a feasibility trial to quit smoking through an on-screen recruitment message presented on the initial onboarding screens within the app. The message asked the following: “Interested in a research study to test the latest version of Smoke Free? If you qualify you could earn up to US $100 and get some paid features for free!” The message included a hyperlink to a screening questionnaire and consent form in the Qualtrics web-based survey platform. Recruitment remained open until the prespecified sample size of 30 participants was met.

Inclusion criteria were the following: aged ≥18 years; downloaded and opened the Smoke Free app; lived in the United States; spoke, wrote, and read English; smoked at least 1 cigarette every day; planned to quit smoking within the next 7 days; and was willing to self-administer a salivary cotinine test remotely. At the end of the screening questionnaire, eligible respondents were asked to consent electronically to participate in the study. Respondents’ eligibility data were automatically transmitted to Smoke Free’s servers to designate each person and consent form in the Qualtrics web-based survey platform. Recruitment remained open until the prespecified sample size of 30 participants was met.

Participants were invited to use the Smoke Free app in whatever way they liked for 28 days after their planned quit date. Therefore, all participants had the same amount of time with the app following their initial quit date. Use was self-directed and did not require any specific amount of time.

We collected self-administered questionnaire data from participants via Qualtrics at baseline and weekly on days 7, 14, 21, and 28 after the participant’s planned quit date. After completing the screening questionnaire, meeting the inclusion criteria, and completing the informed consent in a web-based form, participants were automatically sent an email link to the baseline questionnaire. The 28-day (“1-month”) follow-up assessment was the primary end point. Participants were sent 3 SMS text messages (1 invitation and 2 reminders) and 1 email reminder for the 1-month follow-up survey, using multiple modalities to try to increase response rates.

All participants who completed the 1-month questionnaire, regardless of smoking status, were invited to complete a salivary cotinine test to biochemically verify their smoking status. Within 2 business days of completing the 1-month follow-up questionnaire, participants were mailed a salivary cotinine test kit and test instructions by 2-day priority mail. We sent an email to participants with a link to a Qualtrics survey in which they were asked to upload a photo of the test results within 7 days of receiving the test kit.

Following the end of the saliva testing period, a randomly selected subset of participants was invited to participate in an exploratory 45-minute semistructured interview via Zoom to share their experiences with using the Smoke Free app.

Participants were eligible for up to US $95 in compensation. They received US $10 for completing the baseline questionnaire within 48 hours; US $5 for completing the weekly surveys on days 7, 14, and 21 within 48 hours; US $30 for completing the 28-day survey within 48 hours of being invited (or US $20 for completing it within 7 days); and US $40 for uploading a photo of their saliva test result within 48 hours of receiving the test kit (or US $30 for uploading a photo within 7 days). All study payments were delivered via email at the end of the study through Tango Card [49], a web-based gift card system that allowed participants to choose from a menu of Visa prepaid cards and hundreds of gift cards.

Outcomes
The study’s outcome measures were selected principally to establish the feasibility and acceptability of the intervention and data collection methods for a planned pilot trial. The study had two primary outcome measures, both focused on user engagement: (1) the total number of unique app sessions from enrollment through 28 days after the user’s initial quit date, measured as the number of times the app was opened, and (2) the mean duration of app sessions, in minutes, through day 28 after the quit date. Both outcomes were passively collected by the app. A new session was defined as opening the app after at least 30 minutes of inactivity.

Secondary outcomes focused on the feasibility and acceptability of the gamified app in the domains of (1) user engagement and retention and (2) user satisfaction. Secondary outcomes were the number of unique days with ≥1 app session; proportion entering Inner Dragon; and satisfaction with Smoke Free (“I liked using the Smoke Free app”) and Inner Dragon (“I liked the dragon game”), reported on a 5-point Likert scale ranging from “Not at all” (score=1) to “Extremely” (score=5).
Tertiary outcomes included the following: (1) the proportion of participants who reported abstaining during the past 7 days at the 1-month follow-up assessment (self-reported 7-day point-prevalence abstinence); (2) biochemically verified 7-day point-prevalence abstinence at the 1-month follow-up assessment, obtained from uploaded results for a self-administered salivary cotinine test (Alere iScreen Oral Fluid Device); (3) program adherence, measured as the number of times of using selected core app features (reported a craving, recorded a diary entry, completed a mission, read a tip, and used the chatbot); (4) number of times of using selected game features (breathing exercise, feeding the dragon, feeding the dragon, memory minigame, using the customization menu to change appearance, and reading the dragon instruction guide); (5) motivation to quit at the 1-month follow-up, reported on a 10-point scale ranging from not at all motivated (score=0) to very motivated (score=10); and (6) digital therapeutic alliance, measured from the bonding and confidence subscales of the Mobile Agnew Relationship Measure, reported on a 4-point measure and dichotomized at a threshold of “Agree” or higher [43,44].

Data Analysis
We started by calculating the descriptive statistics of demographic and smoking factors. For the primary aim, descriptive statistics were used to estimate user engagement with the app. A complete case analysis was conducted, including only those who opened the app after completing the screening questionnaire.

To explore the association between the total number of app sessions and the number of sessions with game use, we estimated an unadjusted linear regression of the relationship between the total number of app sessions and the number of sessions with game use. The total number of app sessions is equivalent to the total number of sessions with use of core features, because each session starts on the main dashboard screen that includes certain core features (eg, calculator of money saved by not smoking).

Smoking abstinence was assessed using both complete case analysis and assuming that nonrespondents were still smoking (ie, “missing = smoking”). Imputation methods were not used because of the small sample size.

As this was a feasibility study, the target sample size of 30 participants was based on budget and feasible accrual during the study timeline, rather than on a power analysis. Quantitative analyses were performed using Stata (versions 16.1 and 17.0; Stata Corporation).

Exploratory Qualitative Interviews
To assess the opinions about and use of the Smoke Free app with Inner Dragon, we invited the feasibility study participants to participate in a 45-minute semistructured interview via Zoom. The exploratory interviews occurred following the saliva testing period. Interview respondents received a US $50 gift card of their choice. The interview included topics about overall app experience, Inner Dragon game experience, and study and data collection processes (Multimedia Appendix 2). The qualitative interview data were compiled and summarized into key points.

Ethics Approval
The institutional review board of the University of California, San Francisco, approved study procedures for phase 1 and phase 2 (19-29335).

Results
Game Design and Development
In total, 3 FGDs were conducted via Zoom with 15 participants, each with 3 to 7 participants, between December 4, 2020, and December 11, 2020. Although participants had varied reactions to the prototyped concept, overall, the reaction was positive. Of the 15 FGD participants, 10 (67%) strongly agreed that they would like to use the gamified app frequently. Even when users expressed doubt or dislike, most (12/15, 80%) reported that the game, as presented, would likely encourage them to check the Smoke Free app more frequently, thus supporting the main study hypothesis.

FGD participants gave feedback about several aspects of the game. They tended to be the most enthusiastic about the “dragon park” and the ability to interact with other users and avatars. FGD participants wanted the game to hold them accountable, including by imposing a meaningful penalty for relapse. They preferred game features that facilitated distraction from and coping with cravings over pet care features, such as feeding the dragon, that were not directly connected to the quit attempt. FGD participants wanted the game to progress alongside the dragon’s growth, for example, making leveling up more challenging as the game progresses. They tended to view the customization options as a means of distraction, rather than as motivation to continue engaging with Inner Dragon over time. Feedback from the FGDs and internal testing were incorporated into the final game design.

Feasibility Trial
Sample Characteristics
We screened 86 new Smoke Free app users over 3 days (September 8, 2021, to September 10, 2021) to reach our goal of enrolling 30 eligible, consenting participants for the feasibility trial (Figure 5). Of the 86 screened individuals, 56 (64%) were ineligible, of whom 9% (n=5) lived outside the United States, 11% (n=6) did not smoke daily, 93% (n=52) were not planning to quit within the next 7 days (including n=43, 77% who had quit already), and 14% (n=8) did not agree to complete a saliva test. Of the 30 enrolled participants, 24 (80%) opened Smoke Free after completing the screening questionnaire, thereby completing the onboarding process within Smoke Free. This is the denominator used for most of the analyses. There was 54% (13/24) response rate for the baseline questionnaire, 46% (11/24) response rate for the 1-month follow-up questionnaire, and 45% (5/11) response rate for the saliva test (only those completing the follow-up questionnaire were asked to complete the saliva test).

Table 1 shows the demographic and smoking-related characteristics of enrolled participants. Most (17/24, 71%) identified as female, and median age was 37 years. Most (21/24, 88%) were non-Hispanic White, and 8% (2/24) were
non-Hispanic Black and 4% (1/24) was Hispanic. Among the 18 baseline survey respondents, median income was US $72,500, and there was a range of educational backgrounds, with 39% (5/13) having received a bachelor’s or associate degree, 31% (4/13) having completed some college or technical school, and 15% (2/13) having received at most a high school diploma.

Participants who completed the baseline survey were asked questions about their smoking behavior (Table 1). Participants smoked a median of 15 cigarettes per day and had made a median of 5 past quit attempts. Participants had been smoking for a median of 18.5 years. Approximately 25% (3/12) of the participants reported using electronic nicotine delivery systems in the past 30 days, and 8% (1/12) reported using nicotine replacement therapy in the past 30 days. Most participants (18/24, 75%) set a quit date as within 1 day of enrollment.

Participants also reported about their past video game experience (Table 1). More than three-fourths (19/24, 79%) of the participants played video games at least monthly during the previous year, similar to the general population of smokers [20]. Moreover, 33% (8/24) of the participants played video games daily, and 33% (8/24) played at least weekly but not daily. Among participants who reported playing games monthly, mobile games were the most popular gaming platform (13/24, 54%), followed by web games (10/24, 42%).

Figure 5. Participant flow diagram. *Individuals may be ineligible because of >1 factor.
### Table 1. Descriptive statistics of participants\(^a\).

<table>
<thead>
<tr>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
</tr>
<tr>
<td><strong>Gender (n=24), n (%)</strong></td>
</tr>
<tr>
<td>Female 17 (71)</td>
</tr>
<tr>
<td>Male 6 (25)</td>
</tr>
<tr>
<td>Nonbinary or transgender 1 (4)</td>
</tr>
<tr>
<td><strong>Age group (years; n=24), median (IQR)</strong></td>
</tr>
<tr>
<td>37 (26.5-43.5)</td>
</tr>
<tr>
<td><strong>Household income (US $; n=13), median (IQR)</strong></td>
</tr>
<tr>
<td>70,000 (30,000-87,000)</td>
</tr>
<tr>
<td><strong>Education (n=13), n (%)</strong></td>
</tr>
<tr>
<td>High school diploma or less 2 (15)</td>
</tr>
<tr>
<td>Some college or technical school 4 (31)</td>
</tr>
<tr>
<td>Bachelor’s or associate degree 5 (38)</td>
</tr>
<tr>
<td>Graduate degree 2 (15)</td>
</tr>
<tr>
<td><strong>Racial and ethnic identity (n=24), n (%)</strong></td>
</tr>
<tr>
<td>Non-Hispanic Black or African American 2 (8)</td>
</tr>
<tr>
<td>Non-Hispanic White 21 (88)</td>
</tr>
<tr>
<td>Hispanic 1 (4)</td>
</tr>
<tr>
<td><strong>Smoking-related characteristics</strong></td>
</tr>
<tr>
<td>Cigarettes per day (n=13), median (IQR) 15 (10-20)</td>
</tr>
<tr>
<td>Past quit attempts (n=13), median (IQR) 5 (3-10)</td>
</tr>
<tr>
<td>Years of smoking (n=13), median (IQR) 17 (11-29)</td>
</tr>
<tr>
<td>Used ENDS(^b) in the past 30 days (n=12), n (%) 3 (25)</td>
</tr>
<tr>
<td>Used NRT(^c) in the past 30 days (n=12), n (%) 1 (8)</td>
</tr>
<tr>
<td>Days to quit (n=24), median (IQR) 1 (0-1.5)</td>
</tr>
<tr>
<td><strong>Video gaming experience (n=24), n (%)</strong></td>
</tr>
<tr>
<td>Frequency of playing video games in the past year</td>
</tr>
<tr>
<td>Daily 8 (33)</td>
</tr>
<tr>
<td>At least weekly but not daily 8 (33)</td>
</tr>
<tr>
<td>At least monthly but not weekly 3 (13)</td>
</tr>
<tr>
<td>Less than once monthly 1 (4)</td>
</tr>
<tr>
<td>Not at all 4 (17)</td>
</tr>
<tr>
<td>Platform used at least monthly in the past year</td>
</tr>
<tr>
<td>Mobile games 13 (54)</td>
</tr>
<tr>
<td>Web games 10 (42)</td>
</tr>
<tr>
<td>Console games 4 (17)</td>
</tr>
<tr>
<td>Other games 4 (17)</td>
</tr>
</tbody>
</table>

\(^a\)Some values are missing for questions asked in the baseline questionnaire. Percentages may not add up to 100% because participants selected >1 race or ethnicity.

\(^b\)ENDS: electronic nicotine delivery system.

\(^c\)NRT: nicotine replacement therapy.

#### User Engagement

User engagement with the gamified Smoke Free app was generally high: a mean of 13.8 (SD 13.1; median 8; range 1-46) app sessions per user and an average of 5.8 (SD 10.6; median 1.1; range 0-81.1) minutes per session (Table 2). Overall, 75% (18/24) of the participants entered the Inner Dragon game at least once. The number of sessions decreased rapidly during...
the first week after the participant’s quit date, for sessions overall and those involving the use of Inner Dragon (Figure 6A). The decline in sessions with game use was similar to the decline in sessions overall, based on the change over time in the share of sessions with game use. The high variability in the total number of sessions by user is shown in Figure 6B. The use of the Inner Dragon game was positively associated with the total number of app sessions (correlation 0.57), such that each additional session with the use of Inner Dragon was associated with 3.1 (SE 0.96) more total sessions ($P$=.003; Figure 7A). In addition, the average time a user spent per session decreased with the total number of sessions for that user (correlation $r$=0.39), such that each additional session was associated with 1.25 (SE 0.63) fewer minutes per session on average ($P$=.06; Figure 7B).

We further characterized the variability in the use of different features of the game. Figure 8 shows the use of various pet care activities in the game. Only 44% (8/18) of the participants who tried Inner Dragon used these pet care activities with some frequency.

Table 2. App use during the intervention.

<table>
<thead>
<tr>
<th></th>
<th>Participants, n (%)</th>
<th>Values, mean (SD)</th>
<th>Values, median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App sessions (n=24)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions per user$^a$</td>
<td>24 (100)</td>
<td>13.8 (13.1)</td>
<td>8 (1-46)</td>
</tr>
<tr>
<td>Minutes per session per user$^a$</td>
<td>24 (100)</td>
<td>5.8 (10.6)</td>
<td>1.1 (0-81.1)</td>
</tr>
<tr>
<td>Unique days with ≥1 session$^b$</td>
<td>24 (100)</td>
<td>6.8 (5.9)</td>
<td>4 (1-19)</td>
</tr>
<tr>
<td>Any use of the Inner Dragon game$^b$</td>
<td>18 (75)</td>
<td>—$^c$</td>
<td>—</td>
</tr>
<tr>
<td>Sessions with any game use per user$^d$</td>
<td>24 (100)</td>
<td>2.4 (2.4)</td>
<td>2 (0-10)</td>
</tr>
<tr>
<td><strong>Pet care events$^d,e$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breathing exercise</td>
<td>24 (100)</td>
<td>1.8 (2.1)</td>
<td>1 (0-6)</td>
</tr>
<tr>
<td>Cleaning the dragon</td>
<td>24 (100)</td>
<td>1.6 (1.9)</td>
<td>1 (0-6)</td>
</tr>
<tr>
<td>Feeding the dragon</td>
<td>24 (100)</td>
<td>3.3 (4.6)</td>
<td>0 (0-15)</td>
</tr>
<tr>
<td>Memory game</td>
<td>24 (100)</td>
<td>2.1 (2.8)</td>
<td>1 (0-10)</td>
</tr>
<tr>
<td>Dragon customizations$^d$</td>
<td>24 (100)</td>
<td>7.2 (15.4)</td>
<td>0 (0-61)</td>
</tr>
<tr>
<td>Read the dragon guide$^d$</td>
<td>24 (100)</td>
<td>4 (4.1)</td>
<td>1.5 (0-12)</td>
</tr>
<tr>
<td><strong>Efficacy (7-day abstinence at 1 month)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported, complete cases (n=11)$^b$</td>
<td>5 (45)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Self-reported, missing cases (=smoking; n=24)$^b$</td>
<td>5 (21)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Verified, complete cases (n=11)$^d$</td>
<td>3 (27)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Verified, missing cases (=smoking; n=24)$^d$</td>
<td>3 (13)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Motivation and satisfaction (n=24)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction—“I liked using the Smoke Free app”$^b,f$</td>
<td>11 (46)</td>
<td>4.2 (0.6)</td>
<td>4 (3-5)</td>
</tr>
<tr>
<td>Satisfaction—“I liked the dragon game”$^b,f$</td>
<td>3 (13)</td>
<td>2.0 (1.0)</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td>Motivation to quit (0-10)$^d$</td>
<td>11 (46)</td>
<td>8.8 (1.7)</td>
<td>10 (5-10)</td>
</tr>
</tbody>
</table>

$^a$Primary outcome.
$^b$Secondary outcome.
$^c$Not applicable.
$^d$Tertiary outcome.
$^e$Participants who did not use the pet care events, including those who did not use Inner Dragon at all, were coded as 0.
$^f$Questions were rated on a 5-point Likert scale: extremely (score=5), very (score=4), somewhat (score=3), slightly (score=2), and not at all (score=1).
Figure 6. Total number of sessions. (A) Percentage of users with a session by day. (B) Number of sessions by users.
**Efficacy of the Intervention**

The percentage of self-reported 7-day point-prevalence abstinence at 1 month was 45% (5/11) among complete cases and 21% (5/24) when assuming that nonrespondents were still smoking. Biochemically verified 7-day point-prevalence abstinence was 27% (3/11) among complete cases who were mailed a saliva test kit and 13% (3/24) when assuming that...
nonrespondents were still smoking. Only 1 self-reported abstainer reported having used e-cigarettes in the past 30 days, but they did not complete the saliva test. None of the participants reported having used nicotine replacement therapy in the past 30 days.

The total number of sessions was positively associated with self-reported abstinence in a simple regression analysis. An additional session was associated with increase in abstinence of 1.3 percentage points when assuming that nonrespondents were still smoking \( (P=.04) \) and of 2.7% points with complete cases only \( (P=.01) \). However, the number of sessions with Inner Dragon use and the ratio of game sessions to total sessions were not associated with either measure of abstinence.

### Satisfaction and Motivation

Among the 12 participants who completed the follow-up questionnaire, participants were very satisfied with the gamified Smoke Free app at the 1-month end point (mean score 4.1, SD 0.8 on a 5-point Likert scale). Satisfaction ratings were slightly higher among those who had used Inner Dragon (mean 4.2, SD 0.2; 9/12, 75%) compared with those who had not (mean 3.7, SD 0.3; 3/12, 25%); however, this difference was not statistically significant according to the Wilcoxon rank-sum test \( (P=.21) \).

Only 25% (3/12) of the participants who had used Inner Dragon provided a satisfaction rating for Inner Dragon, because the question was not mandatory; their ratings were substantially lower (mean 2, SD 0.1) than that for Smoke Free overall.

### Exploratory Qualitative Interviews

Qualitative interviews were conducted with 14% (3/22) of the participants who were invited and agreed to participate in the interview. Overall, 67% (2/3) of the interviewed participants had used the Inner Dragon game. In general, the participants liked the design of the app and had no major complaints. Participants found the progress indicators of health improvements as one of the most popular features, and 67% (2/3) of the participants commented about the helpfulness of the chatbot.

However, the participants reported that they had difficulty in understanding the purpose of the game and how to navigate the game. Participants also reported overlooking certain features of the game, such as the memory minigame. They recommended adding a notification to the dashboard that would link the users directly to the game and providing additional tutorials to explain the purpose of the game.

### Adverse Events and Operational Issues

There were no adverse events reported during the trial. However, the study implementation faced technical challenges. In particular, we discovered that the dragon park had not been operational during the feasibility trial. Therefore, users were not able to use this key game feature, and user satisfaction may have been influenced accordingly. We had also planned to include a once-daily pop-up to assess daily smoking status; however, this also was not operational.

### Discussion

#### Principal Findings

Drawing on multiple behavioral theories, we designed a unique, multifaceted digital pet game to promote user engagement and retention in a leading smartphone app for smoking cessation. Participants, who were given access to the gamified app, averaged 13.8 app sessions and 5.8 minutes per session throughout the 28-day trial. User satisfaction with the app was high; users reported an average score of 4.2 on a 5-point scale when asked if they “liked the Smoke Free app.” Regarding smoking abstinence, 46% (5/11) of the participants reported having abstained from smoking during the previous 7 days, and 27% (3/11) of the participants were biochemically verified to have abstained.

The development process demonstrated the feasibility of integrating the Inner Dragon game within the Smoke Free app. The single-arm trial demonstrated the feasibility of performing planned procedures for a pilot randomized controlled trial and provided proof of concept that the Inner Dragon game might be able to increase the engagement of Smoke Free users. The evaluation found that user engagement with the gamified Smoke Free app was generally high, with considerable variability across participants. Short-term data about app use suggested that users who used the game more frequently were also more likely to use core app features more frequently. A randomized controlled trial could reveal whether this is a causal relationship. Participants in the feasibility evaluation liked the gamified app overall; however, we did not collect sufficient data to gauge their satisfaction with the Inner Dragon game.

Our study makes 2 key contributions to the literature about serious games and gamification for smoking cessation. First, although several existing games such as Tobbstop and Quittr use similar frameworks and game features to promote engagement and smoking cessation, our use of a digital pet is novel for smoking cessation apps. Second, we use the game in a unique context in which it is integrated into an app with 6 million downloads so far and available at a scale that is rarely achieved by serious games.

#### Comparison With Previous Studies

Our study of Inner Dragon contributes to the developing literature about the feasibility, engagement, and effectiveness of gamified smoking cessation apps. A small number of such apps have been developed and evaluated, including Crush the Crave [23,24], QuitIT [26,27], Quittr [28,29], and Tobbstop [30,31]. Overall, the literature highlights low user engagement for gamified smoking cessation apps. Crush the Crave participants reported an average rating of 2.4 out of 5 for their frequency of using the game [23]. QuitIT users also had low user engagement, with only 40% of participants reporting that they played the game [27]. User engagement was high among the Quittr users with access to the Tappy Town game (mean of 58.4 min of using the app over a 28-day trial) [29] and the Tobbstop app in which 65.9% of participants successfully used the app [31]. In comparison, 75% (18/24) of our study’s participants used the Inner Dragon game in the Smoke Free app.
at least once throughout the 28-day trial, resulting in relatively high user engagement compared with other apps.

A small number of randomized trials have been conducted to evaluate the efficacy of gamified apps to improve smoking cessation. Although participants have generally reported high satisfaction with the gamified app, most of the apps have not demonstrated improved smoking abstinence rates [23,27,29,31]. Our feasibility trial found a point-prevalence abstinence that exceeded those reported for most other gamified smoking cessation apps (eg, 4% for Quittr, 14% for Crush the Crave, and 39% for Tobbstop); however, the gamified Smoke Free app will need to be tested in a randomized trial with a control group.

Future Directions
The feasibility evaluation produced several important findings with implications for the planned pilot randomized controlled trial and for future studies more generally. Several lessons relate to the game design. First, users reported overlooking the game entirely or not remembering to open it. To increase the visibility of the game, we subsequently included a message at the top of the main dashboard that linked to the game. A message linking to the game was also added under the support tab, where many other resources, such as the chatbot, were available. Second, to address participants’ misunderstandings regarding the purpose of the game and the availability of game features, we created and embedded a promotional video that plays when the user opens the Inner Dragon screen for the first time. The video describes the purpose of the game and its relationship with the user’s quit attempt and highlights the key features of the game, such as the breathing exercise. We also added and expanded on tutorials to clarify the availability and purpose of certain features. Third, we provided tight integration between Inner Dragon and the core features of Smoke Free through the use of more salient messaging outside the game and navigation buttons to the game. In particular, we added messages on nongame screens to inform participants when they earned experience points for completing core activities. For example, upon logging a diary entry, a message popped up to inform the user that they had won experience points that can be claimed in Inner Dragon, along with a button to navigate directly to Inner Dragon.

The feasibility trial also provided lessons for improving the study design of the planned pilot trial. The low response rates pointed toward a need to send reminder messages regarding questionnaires via SMS text message rather than via email. There was also a need to devise procedures for monitoring which app features were active.

Study Limitations
The study has important limitations. First, the small sample size limited the precision of estimates for main outcomes. Second, and relatedly, low response rates and the potential for selection bias made it difficult to assess certain outcomes, such as user satisfaction. However, our experience allowed us to formulate new procedures to improve participant retention. In response, we have planned a large pilot trial with 500 participants to mitigate this small sample size limitation. Finally, the follow-up period was short, and a long follow-up was implemented in the next phase of the project. Several game features, such as the dragon evolutions, converting experience points into gifts, and social interactions in the dragon park, provide the basis for promoting long-term engagement with the app, and this will require further design and development.

Conclusions
In summary, this trial showed that it is feasible to integrate a gamification component into an existing smartphone app for smoking cessation. Moreover, we developed a set of study procedures that could feasibly be used to evaluate the game intervention. Digital gamification is a promising strategy that merits further attention from smoking cessation researchers.

Acknowledgments
The authors would like to thank Keith McCullough and his team at Workinman Interactive, LLC, which contributed to the game design and provided artwork. Adam MacLean, Elena McGahey, Giselle Edwards, Pallavi Chaudhary, and Janardan Devkota provided valuable research assistance. This study was supported by funding from the National Cancer Institute (R21 CA238301); National Institute of Aging (P30 AG012839); Center on the Economics and Demography of Aging at the University of California, Berkeley; and Hellman Fellows Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US government.

Data Availability
The survey data sets generated and analyzed during this study are not publicly available but are available from the corresponding author upon reasonable request. The app use data generated and analyzed during this study are not publicly available owing to a third-party agreement between the University of California, San Francisco, and 23 Limited (the developer of Smoke Free) that prohibits data sharing.

Conflicts of Interest
DC is the Founder and chief executive officer at 23 Limited, the software development company that owns and operates Smoke Free. EW is the chief technical officer at 23 Limited. CL was a Visiting Researcher at Google during the publication of this study. LA receives royalties for the sale of Text2Quit, a quit smoking program. JT reports membership on the scientific advisory board of MindCotine, Inc, which offers a smoking cessation program. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflicts of interest policies. All other authors declare no other conflicts of interest.
Multimedia Appendix 1
Focus group discussion guide.

[PDF File (Adobe PDF File), 93 KB - games_v11i1e46602_app1.pdf]

Multimedia Appendix 2
Exploratory interview guide.

[PDF File (Adobe PDF File), 88 KB - games_v11i1e46602_app2.pdf]

References


24. Baskerville NB, Struijk LL, Dash D. Crush the crave: development and formative evaluation of a smartphone app for smoking cessation. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e52 [FREE Full text] [doi: 10.2196/mhealth.9011] [Medline: 29500157]


29. Bindoff I, Ling TR, Gee P, Geelan B, Ferguson SG, Peterson GM. Effects of a mobile app called Quittr, which utilizes premium currency and games features, on improving engagement with smoking cessation intervention: pilot randomized controlled trial. JMIR Serious Games 2020 Dec 14;8(4):e23734 [FREE Full text] [doi: 10.2196/23734] [Medline: 33315016]


32. Rajani NB, Mastellos N, Filippidis FT. Impact of gamification on the self-efficacy and motivation to quit of smokers: Observational study of two gamified smoking cessation mobile apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [FREE Full text] [doi: 10.2196/jmir.27290] [Medline: 33904824]


49. Tango card IRG. Tango Card Inc. URL: https://www.tangocard.com/ [accessed 2021-08-20]

Abbreviations

FGD: focus group discussion
mHealth: mobile health
A Serious Game to Self-Regulate Heart Rate Variability as a Technique to Manage Arousal Level Through Cardiorespiratory Biofeedback: Development and Pilot Evaluation Study

Tony Estrella\textsuperscript{1,2}, MSc; Carla Alfonso\textsuperscript{1,2}, MSc; Juan Ramos-Castro\textsuperscript{3}, PhD; Aitor Alsina\textsuperscript{4}, PhD; Lluis Capdevila\textsuperscript{1,2}, PhD

\textsuperscript{1}Laboratory of Sport Psychology, Department of Basic Psychology, Universitat Autònoma de Barcelona, Barcelona, Spain
\textsuperscript{2}Sport Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain
\textsuperscript{3}Group of Biomedical and Electronic Instrumentation, Department of Electronic Engineering, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
\textsuperscript{4}Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Barcelona, Spain

Corresponding Author:
Lluis Capdevila, PhD
Sport Research Institute
Universitat Autònoma de Barcelona
Edifici N, Planta 1
Barcelona, 08193
Spain
Phone: 34 93 581 3329
Email: lluis.capdevila@uab.cat

Abstract

Background: Heart rate variability biofeedback (HRVB) is an established intervention for increasing heart rate variability (HRV) in the clinical context. Using this technique, participants become aware of their HRV through real-time feedback and can self-regulate it.

Objective: The aim of this study was 2-fold: first, to develop a serious game that applies the HRVB technique to teach participants to self-regulate HRV and, second, to test the app with participants in a pilot study.

Methods: An HRVB app called the FitLab Game was developed for this study. To play the game, users must move the main character up and down the screen, avoiding collisions with obstacles. The wavelength that users must follow to avoid these obstacles is based on the user’s basal heart rate and changes in instantaneous heart rate. To test the FitLab Game, a total of 16 participants (mean age 23, SD 0.69 years) were divided into a control group (n=8) and an experimental group (n=8). A 2 × 2 factorial design was used in each session. Participants in the experimental condition were trained in breathing techniques.

Results: Changes in the frequency and time domain parameters of HRV and the game’s performance features were evaluated. Significant changes in the average RR intervals and root mean square of differences between adjacent RR intervals (RMSSD) were found between the groups (\(P=0.02\) and \(P=0.04\), respectively). Regarding performance, both groups showed a tendency to increase the evaluated outcomes from baseline to the test condition.

Conclusions: The results may indicate that playing different levels leads to an improvement in the game’s final score by repeated training. The tendency of changes in HRV may reflect a higher activation of the mental system of attention and control in the experimental group versus the control group. In this context, learning simple, voluntary strategies through a serious game can aid the improvement of self-control and arousal management. The FitLab Game appears to be a promising serious game owing to its ease of use, high engagement, and enjoyability provided by the instantaneous feedback.

\textit{(JMIR Serious Games 2023;11:e46351)} doi:10.2196/46351

KEYWORDS
serious game; heart rate variability; biofeedback; mobile health; mHealth; app; mobile phone
Introduction

Background

Heart rate variability (HRV) represents the temporal variation between successive heartbeats (RR interval), and it is measured in milliseconds [1]. HRV is a biomarker of cardiovascular [2] and mental [3] health, reflecting the balance between the parasympathetic and sympathetic branches of the autonomic nervous system (ANS) [4]. HRV has been considered as an objective and noninvasive health indicator, with higher resting HRV being associated with lower cardiovascular risk [5], reduction of negative emotions through acute stress [6], flexible and adaptive response to environmental demands [7], and improved fitness [8]. This reflects the numerous applications of HRV in various fields.

HRV biofeedback (HRVB) is an established intervention that increases HRV in the clinical context and has proven to be effective as a treatment for anxiety and depression [9,10]. With this technique, participants become aware of their HRV through real-time feedback and can self-regulate it, restoring its homeostasis [11]. HRVB is based on the physiological phenomenon of respiratory sinus arrhythmia, which is defined as the variation in heart rate (HR) caused by respiration [12]. The analysis of respiratory sinus arrhythmia shows that the instantaneous HR (iHR) increases during inhalation and decreases during exhalation. Thus, the original HRVB procedure proposed by its promoters emphasizes the importance of breathing training to maximize HRV and achieve the maximum benefits of this technique [12]. In a recent systematic review [13], a series of methodological guidelines were proposed to improve the effectiveness of HRVB interventions. This review highlighted the importance of the breathing protocol as the core of the HRVB interventions, concluding that concrete aspects such as the inhalation, holding, or exhalation ratio should be explained in detail to the participants who apply the technique. Furthermore, HRVB can help establish a connection between the mind and body through voluntarily controlled physiological behaviors, such as breathing, facilitating the learning process and self-control.

HRVB is becoming more widely used owing to its ease of application and high cost-benefit potential, particularly because of the current accessibility of health technologies through mobile health devices [14,15]. The growth of mobile health technology has allowed the development of game-based mobile apps, which are known to increase the transfer of learning and motivation toward practice [16]. The use of games either in health or educational interventions allows the participants to immerse themselves in the activity that is required by the game, promoting a flow state [17,18]. Similarly, other studies have shown that when playing serious games, the student’s role changes from a passive receiver of information to an active entity interacting with the game, facilitating the learning process [19,20]. Serious games, rather than being geared toward entertainment, are aimed at educating, training, and informing the player about a specific topic [21]. They are being used in various health domains [22-24]. For instance, in the field of HRVB, they can be a way to teach participants to control HRV through breathing techniques in an interactive environment.

Objectives

With the information provided in the Background section, the aim of this study was 2-fold: first, to develop a serious game that applies the HRVB technique to teach participants to self-regulate HRV and, second, to test the app with participants in a pilot study.

Methods

This study has two phases: (1) the development of a serious game biofeedback app called the FitLab Game and (2) the testing of the app with participants in a pilot study.

Phase 1: App Development

Visual Environment and Game Explanation

General Design

In accordance with definition of serious games by Djaouti et al [25], the FitLab Game is meant for physical and mental training in the health care sector and intended for the general public. The overall goal of the game is to help the user self-regulate their HRV by providing biofeedback in an attempt to follow or mimic a preset iHR waveform. The architecture and design of the game allows the incorporation of new real iHR curves (RR) to imitate and can be adapted to different levels of complexity in the different existing scenarios (levels). These real recordings can be chosen by the player or a professional accompanying the player. Depending on the level and the preset iHR waveform chosen, the game will lead to a state of relaxation (increasing cardiac variability) or a state of activation in the participant (decreasing cardiac variability), according to the needs demanded by each particular situation. Different gamification techniques were used in its design to increase the players’ engagement with the tasks. More general techniques include having a goal to aim toward, a set of rules to follow, a challenge to complete, and real-time feedback to tell the player how he is doing and how close he is to completing the goal. More specific gamification techniques were added to further engage the player, including Bonus points, increasing challenge, and performance graphs.

Menus

The game’s menus were designed in landscape mode and presented in Spanish. The first screen that appears when opening the app is the main screen, which allows users to choose the sensor used to input data: either the device’s camera (“Cámara de video”) or a HR band (“Banda Pectoral”). The latter was used in this study. Once connected to the sensor, the app allows access to the game’s mode, where the player can select between the option “Jugar” (“Play”), in which the user overcomes different levels in an orderly manner so that the difficulty of the game gradually increases, or the option “Práctica” (“Practice”), in which the user can activate an independent session. The tab “Opciones” (“Options”) refers to the game’s settings (Figure 1).
Figure 1. Screenshots of the game’s menus. Game’s transition from the main screen to the screen that appears when you select “Banda pectoral” (“Chest Band”) and which allows to connect a heart rate sensor to the game. On the right, the third menu shows where the player can select “Jugar” (“Play”) to play.

Screens

The FitLab Game has 9 levels distributed in 3 visually different worlds (Figure 2). Each world has 3 scenarios or levels of difficulty that the user must overcome in a set order to be able to access the next one when in “Jugar” (“Play”) mode. To overcome a scenario, the user must move the main character up and down the screen to avoid colliding with obstacles that keep appearing. Touching an obstacle results in the loss of a life, and losing all lives means the game is over and needs to be restarted. The movement of the main character reflects cardiac variability and can be controlled by different voluntary strategies such as breathing.

The upper and lower limits and the obstacles in each scenario of the game are not placed randomly. They follow a specific waveform determined by previously recorded RR intervals, corresponding to a real recording. Each scenario represents a real RR series transformed and visualized as iHR, with different difficulty to imitate depending on the particular and distinct RR series (represented as a curve of iHR values) on which each scenario is based. Before starting the game, the data were calibrated to establish an initial HR range for the user (based on iHR from RR intervals). An algorithm was created specifically for the game, which takes this initial HR and displays it on a “grid” (invisible to the user) where each box is 1 second wide per 1 second high. With this algorithm, the screen is configured to place the character in the central area vertically, and the width of the screen is adjusted such that it is proportional to the range of variability presented by the user. In addition, the upper and lower limits and the obstacles, based on the previously recorded RR series, are placed above and below the defined waveform, leaving a space for the character to pass in between (Figure 3).

As a result of the initial calibration, all users start the game under similar conditions, while also having the level of difficulty of each scenario of the game adjusted to their own HR values. Overall, the configuration of each game is a dynamic and individual process, setting it apart from “one-size-fits-all” games. The individually tailored nature of the FitLab Game was designed to increase motivation and the learning process [26].

Figure 3 illustrates the algorithm that defines the ideal pathway of the character. The blue line represents the waveform defined by the prerecorded HR curve (from a specific RR series), and the dark gray boxes represent the upper and lower limits of this curve, where the obstacles would be positioned. The right side of Figure 3 shows an example of how this algorithm would be seen in the game (see a dynamic example in Multimedia Appendix 1). Once the obstacles are placed, the objects and landscape of the scene are selected. For each world, there is a bank of images that can be used (mountains, animals, houses, etc) and placed by distributing them according to their size. Larger objects form the background plane, such as the mountains. Medium objects, such as houses, gradually fill the spaces that the first type of images could not fill because of their size, and small objects, such as birds, trees, or other details, are placed in the spaces that are missing to be filled in, which in Figure 3 are the dark gray boxes. Notably, the placement of obstacles is not exhaustive, meaning that it is not intended to fill all the space reserved for obstacles; otherwise, the screen will be saturated with objects. The goal is to place objects in the most harmonious way possible but clearly marking the path that the main character should follow. When the user reaches the end of the level or loses a life owing to touching obstacles, a final page appears. On this screen (Figure 4), a “Level Completed” message is shown, together with the total points collected, and the average stress index and average HR of the player and of all participants who have played the game is also shown.
Figure 2. Screenshots of worlds and scenarios of the game: “Campo” (“Rural area”), “Mar” (“Sea”), and “Espacio” (“Space”).

Figure 3. Example of wave based on an instantaneous heart rate curve from a particular RR interval series.

Figure 4. Example of the final page, appearing after a word of the FitLab Game is played.
**Data Collection and Bluetooth Connection**

Before communication occurs between the sensor (HR sensor) and the FitLab Game, a connection must be established between the entities. Such a connection is established via Bluetooth and is known as pairing. In this type of connection, one of the devices acts as a host or receiver of data and the other acts as a client or sender. The device acting as a receiver can establish connections with multiple sending devices, but a sending device can only connect to a single receiver. Currently, the present app allows only 1 sender to be connected to the device because each game is calibrated to a user’s iHR.

It should be noted that the iHR reading device used for the FitLab Game could be any, but it must offer a Bluetooth low energy connection so that the app can obtain the data. This version of Bluetooth can define specific data formats using the variation of each iHR beat with respect to the previous one. This app container was implemented using a cross-platform environment for writing applications for Android and iOS operating systems called Titanium SDK (TiDev, Inc).

The game engine is written in JavaScript and is based on a constant loop where, at each step, the current position of the character is evaluated with respect to its environment (screen boundaries, obstacles, and prizes) and the input of the user (iHR data; Figure 5). The game engine, menu screens, and navigation are written in web languages compatible with any browser. Finally, both operating systems (Android and iOS) allow the introduction of web components within the app to reproduce web code (ie, HTML+CSS+JavaScript) as if it were part of the same app. The current app takes advantage of this feature to separate the game from the app’s container, which allows communication via Bluetooth with other devices. When the container receives a command via Bluetooth, the app container code sends the corresponding interpretation of this command to the game part. For example, when instantaneous HR data are received over Bluetooth, the app container unpacks the received frame, obtains specific iHR data, and sends it to the game, which interprets it and moves the character. Figure 5 shows a game flow representation, where the user wearing a chest band sends iHR signals via Bluetooth and the game algorithm evaluates iHR, time playing, Bonus achieved, and position on a constant loop.

**Figure 5.** Game flow representation. iHR: instantaneous heart rate.
connection has been successfully established or if an error has occurred. Once the connection is established, the application takes the user to the game’s playing menu.

**Data Processing**

For each game carried out in the app, iHR data are collected together with other metadata, including the start and end times of the game, a user’s identifier code, an ID assigned to the session, the level of the game played, the Bonus achieved, and the iHR calibration. All these data are packaged, saved in the container of the app, and can later be exported for analysis. Multiple user sessions can be stored locally, that is, in the application, but an internet connection is required to export the data to a server. Note that the game does not need an internet connection to work, and a wireless Bluetooth connection between the sender and the receiver in proximity is sufficient. For research purposes, the sessions that are temporarily saved on the device are accessible from a configuration menu by an application manager. To export data, the manager needs to select the desired sessions and indicate an email address to which the anonymized data will be sent, that is, identified by user codes and never by names or other identifying data.

**Phase 2: Pilot Study**

**Study Sample**

A total of 16 participants (9 men and 7 women; mean age 23, 0.69 years) were divided into 2 groups: a control group (4 men and 4 women; mean age 23.25, SD 0.99 years) and an experimental group (5 men and 3 women; mean age 22.75, SD 1.01 years). All participants were aged >18 years and did not present any cardiovascular disorder or take any medication that could affect HRV. The descriptive statistics of the participants are presented in Table 1. All participants, selected via a convenience sample of university students, were volunteers, and all participants provided written informed consent.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control group (n=8)</th>
<th>Intervention group (n=8)</th>
<th>All participants (N=16)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (woman), n (%)</td>
<td>4 (50)</td>
<td>3 (38)</td>
<td>7 (44)</td>
<td>.61</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>23.25 (2.82)</td>
<td>22.75 (2.87)</td>
<td>23 (2.76)</td>
<td>.73</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>170 (10.53)</td>
<td>173.88 (8.34)</td>
<td>171.94 (9.39)</td>
<td>.43</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>66.5 (13.86)</td>
<td>66 (9.41)</td>
<td>66.25 (11.45)</td>
<td>.93</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>22.84 (3.38)</td>
<td>21.78 (2.50)</td>
<td>22.31 (2.92)</td>
<td>.49</td>
</tr>
</tbody>
</table>

**Material and Instruments**

**Control Measures Self-Report**

Before attending the session, participants were recommended, by email, to avoid taking nonessential drugs (up to 24 hours before the session) as well as caffeine, smoking, or any other psychostimulant (up to 2 hours before); alcohol (10 hours before); or eating in general (1 hour before). They were also asked to avoid high-intensity physical activity or any unusual exercise (20 hours before), to sleep for at least 6 hours, and to wear comfortable clothes to the laboratory. A brief questionnaire was administered immediately before the laboratory sessions to record these conditions. The participants were also asked to report their weight and height. In addition, the BMI was computed based on the participants’ reported height and weight. This measure was used as a control variable between the groups because of its possible influence on HR [27].

**Polar Band H10**

An H10 cardiac chest band (Polar Electro) was used to record the RR interval signal (transformed to iHR), as validated previously [28]. The signal was sent and collected to the FitLab Game App (iOS version 1.0, build 1; Health&SportLab SL), which was active on an iPad Air 2 (version 12.1.4).

**Questionnaire of Satisfaction With FitLab Game**

An ad hoc satisfaction questionnaire based on the proposals of different works [29-31] was tailored to our research purpose. The questionnaire was created to collect participants’ information about the following factors (see item number in brackets): (1) usefulness, (2) ease, (3) enjoyment of the game, (4) personal ability, (5) intention to use, (6) clarity of the goal of the game, (7) controllability of the game, (8) strategic approach, (9, 10, and 11) flow, (12) feedback, (13) educational aspect, (14 and 15) motivation, and (16) design aspects. Each factor was represented by a single item, except for flow and motivation, which consisted of 3 and 2, respectively. The participants responded to the questionnaire using a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Mean scores >3 can be considered as positive, 3 as neutral, and <3 as negative. Items related to the techniques used to modulate HR were not included in the control group. The average total score was computed based on participants’ responses, providing a general satisfaction score.

**Procedure**

The study consisted of 1 individual session per participant, which lasted for approximately 35 minutes and took place in the Basic Psychology Laboratory of the Universitat Autònoma de Barcelona. The day before this session, each participant was provided a list of recommendations, based on the Control Measures Self-report, to record variables that may affect HRV, including physical activity and intake of caffeine, psychostimulants, alcohol, drugs, and food. At the beginning of the session, the H10 cardiac band was placed on the participant’s chest so that they could become familiarized with the sensation of the band. A few drops of conductive gel were applied to the band to facilitate good contact with the skin. Once the band was placed, the participant answered the Control
Measures Self-report questionnaire. The participants were then introduced to the FitLab Game for the first time. Play time was divided into 3 parts, with the first and third parts being recorded for the study (Figure 6). In part 1, the baseline phase, each participant played 2 levels of the game to familiarize themselves with the functioning of the app, without time constraints. In part 2, the practice phase, participants practiced on a different level for 6 minutes, or a maximum of 3 levels. Finally, in part 3, the test phase, each participant played the same 2 levels as in part 1.

Figure 6. Example of screens of the game that participants played. Testing consisted of 3 parts (part 1: baseline phase, part 2: practice phase, and part 3: test phase).

A few techniques to consciously control iHR (and thus manage HRV) were explained to the participants in the experimental group but not to those in the control group. To increase HRV (and reduce iHR), it was recommended to either breathe at a pace where the exhalation was longer than the inhalation [32-34] or to hold one’s breath [35]. To reduce HRV (and increase iHR), participants were instructed to slightly hyperventilate [36], to perform muscle movements [37], or to swallow saliva voluntarily all at once [38].

For each level and session that the participants played, the FitLab Game calibrated a baseline iHR, which allowed the game’s character to be placed in the center of the screen, in such a way that the starting point and the difficulty of the iHR series were equal for all participants. Then, participants had to follow the same pre-established route corresponding to each level and session, collecting apples (Bonus), which marked the correct route, and avoiding other elements such as unhealthy foods, tobacco, or the upper and lower limits marked by the floor and ceiling of the screen. To follow the route (or avoid certain elements), the main character had to move up or down the screen, which was controlled by changes in iHR. Participants in the control group were given no instructions on how to do so, whereas participants in the experimental group were explained the techniques mentioned in the paragraph above. All participants started the game with 20 lives, and the game ended when they ran out of lives or when they reached the end of the level without losing all these lives.

Data Analysis

Descriptive statistics are reported as means and SDs. A $2 \times 2$ multivariate analysis of variance following a general linear model was used to analyze the differences between the baseline and test situations. The baseline values were calculated as the average of the 2 games played in part 1, and the test values were calculated as the average of the 2 games played in part 3 (Figure 6). This statistical analysis was carried out to analyze the baseline versus test data obtained from HRV parameters, as well as time and Bonus, all separately. Statistical analysis was performed using SPSS (version 28.0; IBM Corp).

The raw RR interval values recorded by the Polar H10 cardiac chest band were processed to remove artifacts owing to false positive or false negative detections; afterward, the RR intervals were filtered, and the corresponding iHR values were shown in real time on the screen when the participant was playing. The same processed RR interval values were saved and analyzed to calculate the HRV parameters. A time domain HRV analysis was performed to calculate average RR intervals (RR mean), SD of RR intervals, root mean square of differences between adjacent RR intervals (RMSSD), and percentage of consecutive RR intervals that differ by >50 ms between them. A frequency domain analysis was used to calculate the low-frequency band (low frequency [LF]; 0.04-0.15 Hz) and high-frequency band (high frequency [HF]; 0.15-0.4 Hz). HRV analysis followed the guidelines recommended by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). HRV analysis was performed using MATLAB (R2021 update 3 for 64 bits Windows; MathWorks).

Ethics Approval

This study was conducted according to the local ethics committee for human experimentation (protocol code CEEAH-5745).
Results

Change in Gaming Performance (Pilot Study)
To analyze the participants’ performance in the game, the change in the number of apples collected (Bonus) and the total time played in seconds (Time) were considered, as presented in Table 2. No significant changes were found in the baseline versus test situation when comparing the experimental and control groups. However, there was a tendency for both groups to have increased Bonus and Time from the baseline to the test condition.

Table 2. Game outcomes in baseline and test phases in the control and experimental groups.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Baseline, mean (SD)</th>
<th>Test, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bonus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>16.75 (10.91)</td>
<td>18.06 (9.54)</td>
</tr>
<tr>
<td>Experimental</td>
<td>15.75 (5.34)</td>
<td>20 (10.15)</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>91.08 (26.79)</td>
<td>109.05 (58.17)</td>
</tr>
<tr>
<td>Experimental</td>
<td>91.07 (26.79)</td>
<td>116.67 (52.31)</td>
</tr>
</tbody>
</table>

Participants’ Satisfaction
The satisfaction questionnaire was completed by 13 (81%) of the 16 participants. Overall, participants expressed high satisfaction with the FitLab Game, reporting positive experiences, with a total score of 3.87 (SD 0.42) for the control group and a total score of 4.30 (SD 0.50) for the intervention group. The intervention group answered 16 questions, with 4 specific questions related to the techniques used during the game. Table 3 provides an extended description of the items and their scores for each group.

Table 3. Items and scores of the ad hoc game’s satisfaction questionnaire for the control group (n=5) and intervention group (n=8).a

<table>
<thead>
<tr>
<th>Items</th>
<th>Control group, mean (SD)</th>
<th>Intervention group, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 “I can learn self-control techniques better and faster with games like this.”</td>
<td>N/Ab</td>
<td>4 (0.53)</td>
</tr>
<tr>
<td>2 “Using games like this would be easy for me.”</td>
<td>4.2 (0.45)</td>
<td>4 (1.07)</td>
</tr>
<tr>
<td>3 “Playing to learn techniques is more fun than following a pacer.”</td>
<td>4.6 (0.55)</td>
<td>4.87 (0.35)</td>
</tr>
<tr>
<td>4 “I would be able to play this game even if nobody was there to teach me.”</td>
<td>N/A</td>
<td>3 (0.76)</td>
</tr>
<tr>
<td>5 “In the future, I would play if I wanted to learn how to control my heart rate.”</td>
<td>4.2 (0.84)</td>
<td>4 (0.76)</td>
</tr>
<tr>
<td>6 “The aim of the game was clear the whole time I was playing.”</td>
<td>4 (1.22)</td>
<td>5 (0)</td>
</tr>
<tr>
<td>7 “The instructions for functionality were clear.”</td>
<td>N/A</td>
<td>5 (0)</td>
</tr>
<tr>
<td>8 “While playing, I understood how to achieve points and how to end the game successfully.”</td>
<td>4.2 (1.30)</td>
<td>4.6 (0.52)</td>
</tr>
<tr>
<td>9 “While playing, I was only thinking about the game.”</td>
<td>3.6 (1.14)</td>
<td>4.75 (0.46)</td>
</tr>
<tr>
<td>10 “While playing, I forgot everything else around me.”</td>
<td>4 (1)</td>
<td>4.5 (0.53)</td>
</tr>
<tr>
<td>11 “While playing, I didn’t realise how much time was passing.”</td>
<td>3.2 (0.84)</td>
<td>4.13 (0.83)</td>
</tr>
<tr>
<td>12 “When I did something in the game, I knew if it was right or wrong.”</td>
<td>3.4 (1.82)</td>
<td>4.25 (1.04)</td>
</tr>
<tr>
<td>13 “By playing I learned how the techniques affect my heart rate, and how I can modify it.”</td>
<td>N/A</td>
<td>4.13 (0.64)</td>
</tr>
<tr>
<td>14 “The game format increased my motivation to continue playing and achieve more bonus points.”</td>
<td>3.4 (0.89)</td>
<td>4.63 (0.52)</td>
</tr>
<tr>
<td>15 “A game that applied different breathing techniques would increase my interest in this topic.”</td>
<td>4 (0.71)</td>
<td>4.63 (0.52)</td>
</tr>
<tr>
<td>16 “The levels are fun and esthetically pleasing.”</td>
<td>4.2 (0.45)</td>
<td>4.38 (0.74)</td>
</tr>
</tbody>
</table>

aTotal score: control group—mean 3.87 (SD 0.42); intervention group—mean 4.30 (SD 0.50).
bN/A: not applicable.

Change in HRV Parameters
When analyzing the changes in HRV before and after the game, the results showed some differences between the experimental and control groups (Table 4). In particular, the experimental group showed a decrease in RR mean from baseline to test, which was different from the increase observed in the control group.
group ($P=.02; \text{Figure 7}$). In contrast, the experimental group showed a decline in RMSSD from baseline to test, which was more pronounced than that in the control group ($P=.04$). No other HRV parameters presented a significant or tendential change.

**Table 4.** Multivariate analysis of variance of heart rate variability variables in baseline and test conditions in the control and experimental groups.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Baseline, mean (SD)</th>
<th>Test, mean (SD)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR mean$^a$</td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Control</td>
<td>800.33 (109.02)</td>
<td>816.89 (106.59)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>1004.52 (148.41)</td>
<td>930.90 (132.45)</td>
<td></td>
</tr>
<tr>
<td>SDRR$^b$</td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>Control</td>
<td>62.55 (24.12)</td>
<td>65.64 (20.16)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>111.35 (33.97)</td>
<td>109.81 (32.44)</td>
<td></td>
</tr>
<tr>
<td>RMSSD$^c$</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Control</td>
<td>43.10 (19.96)</td>
<td>41.60 (20.16)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>93.80 (40.71)</td>
<td>80.07 (31.73)</td>
<td></td>
</tr>
<tr>
<td>pNN50$^d$</td>
<td></td>
<td></td>
<td>.36</td>
</tr>
<tr>
<td>Control</td>
<td>23.83 (17.53)</td>
<td>20.77 (17.25)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>42.90 (18.63)</td>
<td>34.64 (17.52)</td>
<td></td>
</tr>
<tr>
<td>LF$^e$</td>
<td></td>
<td></td>
<td>.39</td>
</tr>
<tr>
<td>Control</td>
<td>3063.58 (2789.61)</td>
<td>2595.49 (2190.80)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>6612.13 (3438.65)</td>
<td>6972.52 (3760.22)</td>
<td></td>
</tr>
<tr>
<td>HF$^f$</td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Control</td>
<td>661.70 (458.82)</td>
<td>739.14 (547.80)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>2985.17 (2311.75)</td>
<td>2046.96 (1594.17)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$RR mean: average RR intervals.
$^b$SDRR: SD of RR intervals.
$^c$RMSSD: root mean square of differences between adjacent RR intervals.
$^d$pNN50: percentage of consecutive RR intervals that differ by >50 ms between them.
$^e$LF: low frequency.
$^f$HF: high frequency.
**Figure 7.** Mean, 95% CI, and distribution of average RR intervals (RR mean) and root mean square of differences between adjacent RR intervals (RMSSD), when comparing baseline and test situations in the experimental and control groups.

---

**Discussion**

**Overview**

The aim of this study was 2-fold: to develop a serious game that teaches the user to self-regulate iHR using HRVB and to carry out a pilot study to test the game with participants. A total of 16 participants took part in the pilot study and were divided into a control group and an experimental group. The results were analyzed in 2 parts: one focused on changes in the performance of the game and the other on fluctuations on HRV parameters. The results related to the usefulness of the game are discussed in the *Game and App Performance* section.

**Game and App Performance**

Analyzed globally, the gaming performance of all participants was in line with the results of another gaming study [39], in which the participants played a biofeedback game 6 times and both the control group that received simulated feedback and the experimental group that received physiological feedback improved their gaming performance. These results indicate that playing a given game, in itself, leads to improvement in its final score through repeated training or habituation. That is, both groups got to experiment and learn how to move the game’s character. Regarding the differences in performance between groups, the results showed that the experimental group had a greater improvement in all indicators compared with the control group, even though the differences were not significant. The experimental group showed a greater improvement in Bonus score (increase of 4.25 Bonus points in the test with respect to the baseline), compared with the control group (increase of 1.31 Bonus points in the test; Table 2). The same trend was observed in the total playing time, where the experimental group kept playing 7.6 seconds longer than the control group, but these results were not statistically significant (Table 2). All these results are consistent with the idea that the application of specific strategies in the experimental group could have improved the self-regulation of heart variability, more than simply practicing the game (as shown with the control group). However, these results should be interpreted with caution because of the small sample size and the high values of SDs shown in Table 2.

In our study, a few gamification techniques were introduced to increase players’ desire to engage with the tasks. General techniques include having a goal to aim toward, a set of rules to follow, a challenge to complete, and real-time feedback to tell the player how he is doing and how close he is to completing the goal. More specific gamification techniques included Bonus points, increasing challenges, and performance graphs. Reynard et al [40], in a systematic review and meta-analysis of digital interventions for emotion regulation, found that digital game interventions consistently provided stronger evidence compared with other types of digital interventions. In this review, 64% of digital game studies provided feasibility for engagement, implementation, adherence, expectations, and transference to real life. However, they did not explain which elements of the specific game design were responsible for this feasibility. For their part, Maher et al [41] concluded that gamified apps were used substantially longer than nongamified apps. Other studies highlight specific game elements that impact health behavior.
and achieve adherence, similar to digital rewards, such as points, badges, or rewards that we used in our game [42]. Similarly, in the design of our game, we complied with five gamification principles relating to the motivation of the user [43,44]: (1) meaningful purpose, as the goal of the users participating in the study was to learn a technique to improve their health; (2) meaningful choice, as users could control how they achieved their goal of eating the apples on the screen; (3) supporting player archetypes, as in our game, the initial iHR level is individualized based on the user characteristics; (4) feedback, as the user’s actions to control the movement on the screen based on their own cardiac variability affect progress, and it is clearly communicated; and (5) visibility, as the amount of progress made in each level (apples, points, and Bonuses achieved) is visible at all times. Most of these principles are reaffirmed by the users’ responses to the subsequent questionnaire, as discussed in Participants’ Satisfaction section. Thus, the different techniques applied in the present serious game could contribute to behavioral changes, such as increasing players’ engagement with the tasks and adhering to the use of the game. For example, the instant feedback provided by the game may have helped produce high engagement and enjoyment for the participants, as noted by Siriaraya et al [45]. At the same time, the fact that the levels of the game are adapted to each participant’s cardiac recording means that each person plays in a personalized environment, which is thought to increase motivation toward the game [46]. Finally, regarding the app’s functionality, we highlight the possibility of using either Bluetooth devices or photoplethysmography sensors [47], allowing participants to play in different contexts and with different resources.

The serious game created for this study can be used in multiple situations (in the long term). The game provides flexibility to play in new scenarios, depending on the specific training or situation goals set by the player. Furthermore, there is a menu for practice in which the players can play to test different techniques. Therefore, this game supports repetitive practice as an effective training method. In addition, the game uses the Bonus and time measurements to self-track the learning progression. As the players enhance their level of expertise, the Bonus points and time increase accordingly. This mechanism ensures that as the player improves, they are rewarded with higher Bonuses and more time in the level, motivating them.

Participants’ Satisfaction

Participants expressed an overall positive satisfaction with the FitLab Game, with a greater total score in the intervention group (4.30 out of 5) than in the control group (3.87 out of 5). As Table 3 shows, this higher satisfaction in the intervention group is also supported by the fact that all the items that both groups had in common showed a higher score of 4 (out of 5) only for the intervention group. Both groups considered the game to be easy and fun (punctuations >4 for both groups). The goal and strategic approach to achieve Bonuses and continue playing were clear for both groups, with scores >4. Furthermore, both groups expressed positive intention (score >4) to continue playing the FitLab Game and learn how to control their HR in the future. The participants in the intervention group rated the feedback provided by the game more positively (score >4) compared with the control group (score >3). Regarding the experience of flow during the game, although both groups scored >3, the intervention group scored >4 across the 3 items. Similarly, in terms of motivation, the intervention group scored >4 and the control group scored >3. Finally, both groups found the screens visually appealing and balanced in terms of colors, with scores >4.

As shown in Table 3, the techniques provided to the intervention group were found to be useful in establishing control over iHR and were considered important for playing successfully. Thus, the neutral score in the personal ability factor indicates that the instructions based on different techniques for controlling iHR should be considered necessary for success in the game. The gameplay was understood by both the groups, and the game was generally perceived as easy and enjoyable. With regard to flow and motivation, the intervention group reported a more favorable experience with the game. This can be explained by the use of techniques that enable participants to exert voluntary control over the characters, thereby enhancing the game experience through the interactive component of the game. These findings are in line with those of previous studies, indicating that a gamified visualization component results in higher intrinsic motivation [48]. It is worth noting that the participants expressed a strong interest in continuing to play the FitLab Game in the future, reflecting their engagement and willingness to play to learn about cardiac variability control.

On the basis of the results presented, we propose an extensive use of the game with the main goal of helping individuals learn how to self-regulate their cardiac variability to achieve emotional self-control. This could be applied in different contexts, including clinical (eg, in anxiety or stress disorders) and nonclinical populations. This is in line with the results of a systematic review where consistent evidence was found for reduced negative emotional experience provided by digital games [40]. Ideally, these games should be used under the supervision of health professionals to facilitate learning-specific techniques for controlling cardiac variability. This is reaffirmed because the only response <4 in the specific items for the intervention group is for the item: “4 I would be able to play this game even if nobody was there to teach me” (rating of 3 out of 5). Nonetheless, as shown in Table 3, participants reported that the game was easy to understand and play (item 2), indicating that both the control and experimental groups understood the game well.

HRV Analysis and Biofeedback

Regarding HRV, the results of the pilot study showed some differences in the cardiac behavior of the experimental and control groups (Table 4; Figure 7). This suggests that the strategies used by the experimental group to compete in the game improved their control of cardiac variability, which in turn may have enhanced their performance. Owing to its pilot nature and limited number of participants, the precise reasons underlying the relationship between control of cardiac variability and performance cannot be conclusively determined. Nonetheless, it is possible that the results reflect the activation and the control of attention mechanisms in the experimental group. If we consider that the game requires a certain level of
concentration and attention to play, especially if the participants are remembering and applying the techniques that were just learnt, some arousal level is needed, reflected in a reduction in the activity of the parasympathetic nervous system, as shown by a reduction, although not significant, in the experimental group (HF parameter) compared with a slight increase in the control group. This result is consistent with the increase, but not significant, in the LF values in the experimental group compared with the decrease in the control group. There is evidence that higher LF values are related to a greater activation of the sympathetic nervous system [49]. At the same time, a significant reduction in cardiac variability was observed in the experimental group, as shown by the RR mean and RMSSD parameters. The set of previous results could explain the improvement in the control of arousal or activation in the experimental group. In line with this idea, Fuentes-García et al [50] found that in chess games, the player’s level and the difficulty of the game determined the activity of the ANS of the individuals, with higher demands in attentional focus resulting in lower HRV. Thus, in this study, the experimental group may have shown reduced vagal activity and cardiac variability because they were given techniques to control their HRV, requiring greater attention when playing [51]. In that sense, HRV is an indicator that the experimental group was active and applying what had been taught.

An array of scientific papers highlights the potential of incorporating biofeedback into different types of digital game interventions, including virtual and augmented reality, internet therapy, biofeedback and neurofeedback, digital games, and web-based programs. In a systematic review by Reynard et al [40], 39 interventions were identified, 10% of which were based on biofeedback. The effectiveness and impact of incorporating biofeedback into digital games have been tested in a range of topics, from stress reduction to physical rehabilitation [52]. For instance, Knox et al [53] and Wencz et al [54] studied biofeedback-driven digital game interventions for stress levels before the emergence of smartphones and concluded that symptoms of anxiety and depression were significantly alleviated in children. Similarly, Jeričić and Sundstedt [55] showed that visual and gameplay biofeedback in serious games supported emotional regulation skills. Moreover, aside from the impact on the psychophysiology of individuals, digital interventions appear to be accessible and attractive compared with one-to-one clinical therapies [40], particularly for the younger population [56]. Lüddecke and Felnhofer [57] demonstrated that biofeedback interventions increased the motivation and long-term participation of users in health-related interventions. Finally, games that incorporate biofeedback, compared with other digital interventions, seem to provide the most evidence for emotional skills transference to real life [40].

In conclusion, there is a growing body of research that emphasizes the utility of incorporating biofeedback into digital game interventions for its attractiveness and potential to enhance emotional regulation, among other benefits.

Returning to our study, our results seem to contradict most studies using the HRVB technique, which conclude that this technique helps improve HRV [58,59] with the general aim of achieving relaxation. However, this was not the case. These articles used set breathing paces, usually between 4.5 and 6.5 times per minute, as HRVB to induce relaxation [60,61]. For example, on the topic of gaming, Al Osman et al [62] used game-based biofeedback to control the activity of the ANS and showed that HRV increased following a controlled breathing rate. In contrast, the goal of this study was not to increase, maintain, or decrease HRV but to teach participants to control its rises and falls through various techniques, including breathing control, but not at a set pace. Thus, our study is not comparable with most other studies, given that we did not focus on breathing paces but rather on an array of techniques to teach the participants how to increase or decrease their HRV. These techniques include the voluntary production of apneas to suddenly increase cardiac variability [35], the performance of muscle movements to achieve a reduction in HRV (in line with Guidi et al [37]), or voluntary swallowing. This last strategy can produce changes in some HRV parameters in healthy people, such as SD of RR intervals, LF, and HF power [63], and effortful swallowing increases LF power and the LF/HF ratio [64]. On a practical level, in our case, these strategies allowed the player to move the character up and down the game screen because they provide rapid changes in cardiac variability.

**Limitations and Future Directions**

This study had 3 main limitations. The first limitation is the sample size of the pilot study. Even if the results show significance in the mean iHR (proportional to the RR mean) and RMSSD, they should still be considered preliminary. The second limitation is that participants were not monitored before and after the game but only during it. This means that the HRV indicators were collected as the game was being played, thus in an active state and not in a resting situation. Our interest was to test whether there were changes in the HRV parameters and game performance at all, as a pilot study, but in future studies, it would be interesting to observe the influence of our intervention on basal HRV.

Finally, most previous studies using techniques such as HRVB included a sample of clinical patients, unlike this study, which was conducted in healthy university students. For instance, Siepman et al [65] compared the effects of HRVB in both healthy people and individuals classified as depressed, showing that the latter experienced a significantly greater improvement in HRV, compared with healthy volunteers. Similarly, Pyne et al [66] reported that HRVB training reduces symptoms of posttraumatic stress in older people only. These studies highlight that HRV parameters seem to be difficult to alter with HRVB in healthy and younger people than in older or clinical populations, potentially explaining the results of this study, with few to no significant changes in some HRV parameters. In the future, a larger sample size of healthy university students or testing the FitLab Game in a clinical population could provide more consistent results. Finally, it would also be interesting to provide new strategies for the voluntary control of HRV, perhaps further refining the ones we have already developed in this study. It would also be necessary to deepen the swallowing strategy, as Yildiz and Doma [38] and Yildiz [67] concluded that spontaneous saliva swallowing can change some short-term HRV parameters significantly even in healthy people.
Conclusions
This study developed a serious game, called the FitLab Game, that teaches the user to self-regulate the iHR using cardiac variability biofeedback. The game was also tested with the participants in a pilot study. Overall, the FitLab Game appears to be a promising serious game because of its ease of use, high engagement, and enjoyability provided by the instantaneous feedback. The strategies learned by the experimental group to compete in the game, which included the voluntary production of apneas, muscle movements, and voluntary swallowing, improved the self-control of cardiac variability and, in turn, may have enhanced performance. A significant reduction in cardiac variability observed in the experimental group, based on the RR mean and RMSSD parameters, could explain the improvement in the management of arousal or activation compared with the control group. Thus, the application of specific strategies in the experimental group may have improved the self-regulation of HRV more than simply practicing the game.

The usefulness of HRV as a health indicator is well known, and HRV biofeedback is considered an established intervention in the clinical context. Thus, learning simple and voluntary strategies through the developed game can help improve self-control and arousal management. Ultimately, with sufficient training, the game can become a relaxation technique for the general population or clinical populations, or, for example, an activation technique for athletes in certain situations.

Acknowledgments
This work was supported by grants PID2019-107473RB-C21 and PID2019-107473RB-C22 funded by MCIN/AEI/10.13039/501100011033 by the Spanish Government and 2021SGR-00806 funded by the Catalan Government.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Gameplay of the FitLab Game.
[MP4 File (MP4 Video), 20497 KB - games_v11i1e46351_app1.mp4]

References


42. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [Medline: 25654660]


48. 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]


Abbreviations

- **ANS**: autonomic nervous system
- **HF**: high frequency
- **HR**: heart rate
- **HRV**: heart rate variability
- **HRVB**: heart rate variability biofeedback
- **iHR**: instantaneous heart rate
- **LF**: low frequency
- **RMSSD**: root mean square of differences between adjacent RR intervals
- **RR mean**: average RR intervals

Edited by G Eysenbach; submitted 10.02.23; peer-reviewed by J Robertson, K Stawarz; comments to author 17.04.23; revised version received 12.06.23; accepted 18.07.23; published 24.08.23.

Please cite as:

©Tony Estrella, Carla Alfonso, Juan Ramos-Castro, Aitor Alsina, Lluis Capdevila. Originally published in JMIR Serious Games (https://games.jmir.org), 24.08.2023. 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.
Exploring Key Factors Influencing Nursing Students’ Cognitive Load and Willingness to Serve Older Adults: Cross-sectional Descriptive Correlational Study

Pei-Lun Hsieh¹, PhD; Yu-Rung Wang², PhD; Tien-Chi Huang³, PhD

¹Department of Nursing, National Taichung University of Science and Technology, Taichung, Taiwan
²Department of Nursing, Chang Gung University of Science and Technology, Chiayi, Taiwan
³Department of Information Management, National Taichung University of Science and Technology, Taichung, Taiwan

Corresponding Author:
Pei-Lun Hsieh, PhD
Department of Nursing
National Taichung University of Science and Technology
No. 193, Sec. 1, San-Min Rd
Taichung, 40343
Taiwan
Phone: 886 422196854
Email: peilun@nutc.edu.tw

Abstract

Background: Virtual learning environments (VLEs) use a virtual environment to support learning activities. VLEs are commonly used to overcome the temporal and spatial restrictions of learning activities held in conventional face-to-face classrooms. In VLEs, students can participate in learning activities using the internet, and teachers can provide assistive learning tools during the process.

Objective: The purpose of this study was to investigate the relationships among nursing students’ mental load, cognitive load, and affective learning outcomes in terms of their willingness to serve older adults in an interaction-based educational virtual reality (VR) learning environment.

Methods: This study used a cross-sectional method. A total of 130 students participated in interaction-based VR learning and completed related questionnaires. Descriptive and inferential statistics and stepwise regression for data analysis were used.

Results: The research results revealed that in the dimension of willingness to use VR learning materials, perceived usefulness received the highest score (mean 4.42, SD 0.45). In the dimension of nursing ability, students scored the highest in information management and application ability to care for case patients (mean 4.35, SD 0.54). Correlation analysis revealed that cognitive load during learning and willingness to serve older adults were negatively correlated, whereas willingness to use VR learning materials was positively correlated with nursing ability and willingness to serve older adults. Analyzing the regression coefficients of predictor variables revealed that willingness to use VR learning materials (β=.23; t²=2.89, P=.005) and cognitive load during learning (β=−.35; t²=–.4.30, P<.001) were predictive factors of nursing students’ willingness to serve older adults.

Conclusions: This study demonstrated that students’ willingness to use VR learning materials and their cognitive load during learning affected their willingness to care for older adults. Therefore, the components of mental or cognitive load generate inconsistent predictive effects on affective variables and willingness to serve older adults.

(JMIR Serious Games 2023;11:e43203) doi:10.2196/43203

KEYWORDS
immersive virtual reality learning; VR learning; mental effort; mental load; service willingness; older adult; virtual reality; nursing student; professional education; digital learning; older adult population
**Introduction**

As a consequence of population aging in many countries, nurses must be equipped with the ability to care for older adults, perform early assessments, and establish care demands. In nurse cultivation and education, technology learning elements are incorporated. Virtual learning environments (VLEs) use a virtual environment to support learning activities. VLEs are commonly used to overcome the temporal and spatial restrictions of learning activities held in conventional face-to-face classrooms. VLEs are learning environments free from temporal and spatial constraints that use a web-based virtual space to ensure representational fidelity and facilitate learner interactions. VLEs are typically established on servers.

In a VLE scenario, learners can perceive a construction of identity and a sense of presence and copresence. VLEs provide learners with multimedia materials and can be paired with teaching strategies. According to teaching goals, models can be designed to provide learning activities that assist learners in completing learning tasks [1,2]. VLEs have the following characteristics. (1) VLEs are planned information spaces. (2) VLEs are social spaces in which teachers and students can interact through teaching and learning processes. (3) Information and social spaces can be presented using text and 2D or 3D virtual reality (VR) modes. (4) In a VLE, students can participate in learning activities and discuss the content of the virtual learning space with teachers. (5) VLEs can integrate various techniques and teaching methods to innovate teaching. (6) VLEs can be used in distance teaching or in reinforcing classroom-based learning activities [1-5].

With advancing technology, content that cannot be presented using conventional teaching methods has gradually been converted using VR capabilities to enhance student learning. VR offers students the virtual experience of being in an actual environment and is suitable for assisting learning for scholars, students, nursing personnel, and patients [6]. In VLEs, students can participate in learning activities using the internet. During the process, teachers can provide assistance through VR learning tools, such as web-based discussion and exchange, web-based tests, uploaded files, assignment submission, peer discussion, peer assessment, grading, questionnaires, and learning-history tracking, thereby generating rich interactive learning experiences [1,2,7].

Hodgson et al [8] used interactive VR lesson plans to support discussion and teaching, discovering that 65% of study participants reported that using VR for case discussions could increase their learning interests and assist them in preparing for future clinical internship scenarios. For the field of nursing, Chiou et al [9] established a VR learning platform to simulate the clinical scenario medication administration. They stated that students exhibited interest in the learning process and improved self-learning efficacy. Ball [10] applied the VR 360 Photosphere environmental platform experimentally, determining its ability to reduce students’ anxiety about actual clinical environments. Herral et al [11] used 360-degree VR videos to simulate patients with trauma for the teaching of nursing students. The videos were divided into topics such as introducing the medical team, admitting patients into the emergency department, inspecting patients, and monitoring patients’ conditions after they have been stabilized. These 4 scenarios supported the learning of students and medical and nursing personnel. Using the number of interactions noted during video coding, the duration of viewing a video, and the questions and answers raised on the basis of a video, the researchers discovered that students exhibited significant differences in their cognition and learning processes. Hodgson et al [8] developed VR videos on the basis of real-world cases in hospital settings. The VR scenarios involved the interaction between patients and the medical team. Students were immersed in an environment to which they did not usually have access, enabling learning ahead of time. The research results indicated that 65% of the participants reported that discussing VR-based cases increased their learning interests and prepared them for clinical work.

Through a scoping review, Fealy et al [12] explored the effect of using immersive VR in nursing and midwifery education and discovered that using VR for simulated scenario learning provided students with the opportunity for repeated learning, which enhanced students’ care ability and confidence. Foronda et al [6] and Shorey and Ng [13] conducted systematic reviews to investigate the effect of virtual simulation teaching on nursing students, nurses, and nursing education. The results revealed that VR for simulated scenario learning could replace or supplement traditional nursing education, with the VR-based method effectively increasing students’ cognition and knowledge, skills and performance, critical thinking, self-confidence, and learner satisfaction. Students can learn from VR simulation scenarios that enable interaction. In this study, we conducted a literature review of domestic and foreign studies and discovered that few studies have explored the use of interactive VR learning and its effect on nursing students’ nursing ability and willingness to serve older adults. This study explored self-evaluated nursing ability and willingness to care for older adults in nursing students after they used interactive VR to learn; the influencing factors were also examined. The results can be used to enhance nursing students’ nursing ability in terms of older adult care.

The research questions (RQs) of the study are as follows. RQ1: What are nursing students’ perceptions of their adoption of interaction-based VR learning? RQ2: What are the nursing students’ perceptions of factors (eg, willingness to use VR learning materials for older adult assessment, cognitive load, and willingness to serve for older adult care) that affect the adoption of VR educational technologies?

**Methods**

**Research Design and Research Participants**

This study adopted a cross-sectional correlation research design. A science and technology university in Taiwan acted as the recruitment site. Purposive sampling was used, and fourth-year students in the nursing department of a 5-year program were recruited as research participants, with those who had suspended their studies excluded. After the participants interacted with the VR-based older adult care assessment system, they completed...
a questionnaire survey. G*Power software (version 3.1.9.2; Heinrich-Heine-Universität Düsseldorf) was used to estimate the number of research samples. We set the statistical power as .8, type I error ($\alpha$) as .05, and effect size as .3 (moderate effect). We identified 8 independent variables in this study. The calculation results indicated that the study required a sample size of 121; a total of 130 nursing students were recruited.

The research participants were enrolled in an older adult nursing course. In addition to classroom learning, they were instructed to use interactive VR lessons to experience older adult care. The VR lesson plan consisted of 2 units, namely the Ascertain Dementia-8 assessment and Instrumental Activities of Daily Living scale (IADLs). In the VR scenario, the students could interact with several older adults. The students could choose suitable assessment scales according to the older adults’ conditions. For example, if an older adult or primary caregiver reported a decline in the older adult’s memory function, then the student could select the Ascertain Dementia-8 assessment to assess the presence of early-onset dementia. If an older adult had reduced activity functions, then the IADLs could be selected. Using voice recognition and a VR images scale, the students could check the correct answers to scenario-based questions. A full test score of 100 was possible, with 60 representing the passing score. The detailed VR operation process is illustrated in Figure 1.

**Research Instruments**

An anonymous structured questionnaire was used for data collection. All research instruments were applied under the developer’s consent. The scales were described as follows.

**Personal Attributes and Demographic Information**

We composed this part of the questionnaire through referencing the literature. Personal attributes consisted of age, sex, education level, and experiences with using VR to learn.

**Willingness to Use VR Learning Materials for Older Adult Care and Assessment**

We referenced the scale based on the technology acceptance model developed by Davis and Venkatesh [14], revising this scale to consist of perceived usefulness (6 items), perceived ease of use (3 items), attitude intention to use technologies (3 items), and perceived flexibility (3 items) regarding self-assessed VR-based older adult care learning materials. A 5-point Likert scale was used for assessment, comprising the options of strongly disagree, disagree, neutral, agree, and strongly agree.

**Cognitive Load Scale for Using VR-Based Learning Materials for Older Adult Care and Assessment**

Referencing Hwang et al [15] and Paas et al [16], we developed a cognitive load scale consisting of 2 dimensions, namely mental load and mental effort. Mental load was used to assess the perceived burden generated for nursing students when they complete the VR-based older adult care learning tasks. Mental effort was applied to evaluate the degree of cognition and amount of resources required by the nursing students when they complete the VR-based older adult care learning tasks, thereby facilitating the measurement of the perceived difficulty of the learning materials and tasks, the formats in which learning materials were presented, and the students’ feelings about the explanation methods. A 5-point Likert scale was used for
assessment, comprising the options of strongly disagree, disagree, neutral, agree, and strongly agree.

**Nursing Students’ Service Willingness Scale Regarding Older Adult Care**

For the nursing students who used the VR lesson to learn to assess older adults, we used a scale to assess their willingness to serve older adults. The questionnaire design referenced that developed by Hsieh et al [17]. This willingness to serve older adults scale comprised 15 items rated on a 5-point Likert scale, ranging from strongly disagree to strongly agree; the higher the score, the greater the student’s willingness to serve older adults. All questionnaire items were direct items. Experts were invited to modify the items to ensure construct validity.

**Questionnaire Reliability and Validity**

**Validity**

To ensure the content validity and expert validity of the questionnaire, we invited 3 nursing professors in departments related to older adult care and 2 industry experts to review the adequacy of the content of each scale in the questionnaire. We referenced Polit and Beck [18], who suggested a calculation method for content validity indicators in which items scored higher than 3 by experts are retained. After calculating the percentage, we determined that the content validity index of the willingness to use VR learning materials for older adult care and assessment was .98, that of the cognitive load scale for using VR-based older adult care learning materials was .96, and that of the nursing students’ willingness to serve older adults scale was .99. Each scale exhibited favorable content validity.

**Reliability**

Cronbach α coefficient analysis was applied. The Cronbach α of the willingness to use VR learning materials for older adult care and assessment was .96, that of the cognitive load scale was .98, and that of the willingness to serve scale was .97. All scales exhibited favorable internal consistency.

**Ethics Approval**

Ethical approval for the study was obtained from the Human Research Ethics Council (CRREC-110-088). The researcher explained the research purpose and process to the participants. To ensure the protection of their personal rights, written informed consent was obtained prior to data collection.

**Data Processing and Analysis**

This study used the SPSS statistical software (version 23.0; IBM Corp) for data entry and statistical analysis. A P value of <.05 indicated statistical significance. Descriptive statistics involved using the percentage, mean, and SD to present distributions. The independent sample 2-tailed t test, one-way ANOVA, Pearson correlation coefficient, and stepwise regression were used for inferential statistics and analysis.

**Results**

**Descriptive Variables of Research Participants**

Of the 130 participants of this study, most (n=126, 96.9%) were women. Most of the participants were aged 19 years (n=66, 50.8%), followed by those aged 18 (n=62, 47.7%) and 20 (n=2, 1.5%) years. A total of 93.9% (n=122) of the participants had previously used VR, but none had used VR learning materials related to the care of older adults.

**Participants’ Use of VR-Based Learning Materials for Older Adult Care and Assessment and Their Nursing Ability**

Regarding the overall research variables, the total mean score of cognitive load during learning was 2.49, that of willingness to use VR learning materials was 4.23, that of nursing ability was 4.25, that of willingness to serve older adults was 3.66, and that of the satisfaction toward the VR learning model was 4.14. The detailed data are presented in Table 1.
Table 1. Participants’ use of virtual reality (VR) learning materials to assess older adult care and their nursing ability (N=130).

<table>
<thead>
<tr>
<th>Research variable, secondary dimension</th>
<th>Secondary dimension, mean (SD)</th>
<th>Total, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive load during learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental load</td>
<td>2.45 (0.74)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mental effort</td>
<td>2.52 (0.78)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Willingness to use VR learning materials</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>4.42 (0.45)</td>
<td>N/A</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>4.25 (0.61)</td>
<td>N/A</td>
</tr>
<tr>
<td>Willingness to use</td>
<td>3.96 (0.74)</td>
<td>N/A</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4.29 (0.49)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Nursing ability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic and general educational competencies</td>
<td>4.32 (0.50)</td>
<td>N/A</td>
</tr>
<tr>
<td>Information management and application ability</td>
<td>4.35 (0.54)</td>
<td>N/A</td>
</tr>
<tr>
<td>Health care policies, finance, and environmental capability</td>
<td>4.28 (0.62)</td>
<td>N/A</td>
</tr>
<tr>
<td>Care quality, case content safety, and leadership ability</td>
<td>4.18 (0.70)</td>
<td>N/A</td>
</tr>
<tr>
<td>Nursing practice ability</td>
<td>4.26 (0.49)</td>
<td>N/A</td>
</tr>
<tr>
<td>Preventive health care ability</td>
<td>4.22 (0.54)</td>
<td>N/A</td>
</tr>
<tr>
<td>Teamwork ability</td>
<td>4.09 (0.72)</td>
<td>N/A</td>
</tr>
<tr>
<td>Risk and quality management ability</td>
<td>3.26 (0.75)</td>
<td>N/A</td>
</tr>
<tr>
<td>Ethical ability</td>
<td>3.38 (0.65)</td>
<td>N/A</td>
</tr>
<tr>
<td>Willingness to serve older adults</td>
<td>N/A</td>
<td>3.66 (0.40)</td>
</tr>
<tr>
<td>Satisfaction toward the VR learning model</td>
<td>N/A</td>
<td>4.14 (0.52)</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.

**Difference Analysis of Demographic Data and Research Variables**

The independent sample 2-tailed \( t \) test was used to explore whether sex-based differences were present among the 4 variables of cognitive load during learning, willingness to use VR learning materials, nursing ability, and willingness to serve older adults; Levene test with equal variance was used to obtain the \( F \) test values of the 4 aforementioned variables, which were 4.13, .01, .34, and 5.07, respectively. Among them, cognitive load during learning and willingness to serve older adults had \( P \) values of <.05. Next, the \( t \) value and significance were used for data analysis. The \( t \) values of the 4 variables of cognitive load during learning, willingness to use VR learning materials, nursing ability, and willingness to serve older adults were 92, 1.52, 1.23, and –14.53, respectively. Only the \( t \) value of willingness to serve older adults exhibited a significant difference (\( P < .001 \)). According to the mean values listed in Table 2, women’s willingness to serve older adults was significantly higher than that of men (\( P < .001 \)).

Table 2. Sex-based differences in research variables (N=130).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Participants, n</th>
<th>Research variable, mean (SD)</th>
<th>F test</th>
<th>t test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive load during learning</td>
<td>4</td>
<td>126</td>
<td>3.00 (1.15)</td>
<td>2.47 (0.71)</td>
</tr>
<tr>
<td>Willingness to use VR learning materials</td>
<td>4</td>
<td>126</td>
<td>4.58 (0.48)</td>
<td>4.22 (0.47)</td>
</tr>
<tr>
<td>Nursing ability</td>
<td>4</td>
<td>126</td>
<td>4.53 (0.54)</td>
<td>4.24 (0.47)</td>
</tr>
<tr>
<td>Willingness to serve older adults</td>
<td>4</td>
<td>126</td>
<td>3.10 (0.04)</td>
<td>3.68 (0.39)</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.
Correlation Analysis of Cognitive Load During Learning, Willingness to Use VR Learning Material, Nursing Ability, and Willingness to Serve Older Adults

According to the analysis results summarized in Table 3, the correlation coefficient of cognitive load during learning and willingness to serve older adults was −.374 (P<.001), indicating that these 2 variables were significantly and negatively correlated. The correlation coefficient of willingness to use VR learning materials and nursing ability was .693 (P<.001), indicating that these 2 variables were significantly and positively correlated. The correlation coefficient of willingness to use VR learning materials and willingness to serve older adults was .274 (P<.001), indicating that these 2 variables were significantly and positively correlated.

<table>
<thead>
<tr>
<th>Research variable</th>
<th>Cognitive load during learning</th>
<th>Willingness to use VR learning materials</th>
<th>Nursing ability</th>
<th>Willingness to serve older adults</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive load during learning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>1</td>
<td>__b</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>P value</td>
<td>—</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Willingness to use VR learning materials</strong></td>
<td>−.122</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>P value</td>
<td>&gt;.05</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Nursing ability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−.006</td>
<td>.693</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>P value</td>
<td>.94</td>
<td>&lt;.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Willingness to serve older adults</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−.374</td>
<td>.274</td>
<td>.138</td>
<td>1</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&gt;.05</td>
<td>—</td>
</tr>
</tbody>
</table>

*aVR: virtual reality.
*Not applicable.

Regression Analysis of Willingness to Use VR Learning Materials and Cognitive Load During Learning on Willingness to Serve Older Adults

We applied regression analysis, and the model variance analysis results in Table 4 indicated that the overall regression model reached a significant level ($F_2=15.133, P<.001$), indicating that the overall regression model was statistically meaningful. After adjustment, the coefficient of determination was .180. Following the $F$ test, we conducted an analysis of the regression coefficients of the predictor variables. The results revealed that the standardized regression coefficient ($\beta$) of willingness to use VR learning materials was .23 ($t_2=2.89, P=.005$), reaching a significant level; the $\beta$ of cognitive load during learning was −.35 ($t_2=-4.30, P<.001$), also reaching a significant level. The standardized regression coefficient represented the amount of variance explained by the individual predictor variables for the dependent variables.

Table 4. Regression analysis of willingness to use virtual reality (VR) learning materials and cognitive load during learning on willingness to serve older adults. The adjusted $R^2$ was 0.18 ($F_2=15.13, P<.001$).

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized coefficient, B estimation value (SE)</th>
<th>Standardized coefficient beta distribution</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.30 (.32)</td>
<td>.32</td>
<td>10.36 (2,127)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Willingness to use VR learning materials</td>
<td>.19 (.07)</td>
<td>.23</td>
<td>2.89 (2,127)</td>
<td>.005</td>
</tr>
<tr>
<td>Cognitive load during learning</td>
<td>−.19 (.04)</td>
<td>−.35</td>
<td>−4.30 (2,127)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Discussion

Participants’ Use of VR Learning Materials for Assessing Older Adult Care and Nursing Ability Under Personal Attributes

Among the research participants, 96.9% were women, 50.8% were aged 19 years, and 93.9% had previously used VR but not VR learning materials related to the care of older adults. This result indicates that VR has been widely applied in daily life but little older adult care learning materials had incorporated VR, resulting in few opportunities for students to learn and experience this immersive technology. The overall learning satisfaction of the students using VR learning materials to assess older adult care was high, a result that is in line with those of the systematic literature review of Gasteiger et al [19].
Under the dimension of willingness to use VR learning materials, perceived usefulness scored the highest among all aspects. This result is similar to that of related studies [20-22], indicating that using VR to assist learning reinforced the cognitive level of learners and assisted them in their learning. Regarding the nursing ability dimension, information management and application ability to care for case patients scored the highest among all aspects, indicating that through engaging with the interactive VR scenario, learners’ ability to perceive and respond to scenario-based problems was enhanced. The participants could use voice recognition to respond and select the correct care method. The learning process was valuable for the participants’ future case management ability, a result echoing those of other studies [22,23].

**Critical Factors Influencing Willingness to Serve Older Adults**

This study discovered that willingness to use VR learning materials and cognitive load during learning were critical factors influencing willingness to serve older adults. This result indicated that engaging in the interactive VR scenario increased the students’ willingness to use such materials. Additionally, with VR supporting conventional classroom teaching, students may repeatedly practice older adult care scenarios, allowing them to become familiar with the relevant knowledge on a cognitive level. They can engage in learning without worrying about making mistakes and harming older adults. Thus, students’ learning load was reduced, and their problem-solving ability and self-confidence increased. These results are consistent with those of other studies [17,21,27,29-31].

Studies of willingness to serve older adults [10,13,17,24,32,33] have reported that through learning to interact with older adults, nursing students can observe older adults’ verbal and nonverbal cues, with such interaction increasing their willingness to serve and care for older adults. This study used VR learning materials, and the results revealed that through the VR simulation scenario, the opportunities for students to interact with older adults were increased. Learning through VR scenarios can reduce students’ anxiety when communicating with older adults in a real-world context, thereby increasing students’ willingness to serve older adults.

**Limitations**

This study had some limitations. The respondents provided subjective responses, which could sometimes obscure their actual intention, and more male participants should have been included in the study. Moreover, this study was a quantitative study, which increased the difficulty of understanding the nursing students’ willingness to serve older adults.

**Conclusions**

This survey of 130 preregistered nurses explored factors affecting their willingness to serve older adults. The most influential factors were willingness to use VR learning materials and cognitive load during learning, both of which can be enhanced through older adult care–related courses and practical experience. Nurses play a critical role in the care of older adults, and courses on older adult care are thus crucial for developing sufficient professional human resource capacity. We recommend that VR case scenarios about older adult nursing assessment be incorporated into in-service education. Through repeated practice using these scenarios, nurses’ willingness and ability to care for and serve older adults can increase. Longitudinal follow-up studies of recent nursing graduates (preregistered nurses) who had engaged in VR-based older adult care courses must be conducted to elucidate these courses’ effects on nurses’ older adult care competence and willingness to serve older adults.

**Acknowledgments**

We sincerely express our appreciation to the experts in the geriatric nursing education fields and their academic, policy, and workforce knowledge for helping us adapt the questionnaire to the context and for assessing the content validity of the questionnaire.
We are grateful for the funding we received from the Ministry of Education and the Ministry of Science and Technology (110-2511-H-025-007). The funders had no role in the design of the study; in the collection, analyses, and interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**Authors' Contributions**

PLH and YRW was responsible for the study idea and design. PLH were responsible for the drafting of the manuscript. TCH supported the technical detail problems of the virtual learning system. PLH performed the data collection. YRW and PLH performed the data analysis. PLH made critical revisions to the paper for important intellectual content. All authors read and approved the final manuscript.

**Conflicts of Interest**

None declared.

**References**

19. Gasteiger N, van der Veer SN, Wilson P, Dowding D. How, for whom, and in which contexts or conditions augmented and virtual reality training works in upskilling health care workers: realist synthesis. JMIR Serious Games 2022 Feb 14;10(1):e31644 [FREE Full text] [doi: 10.2196/31644] [Medline: 35156931]


Abbreviations

| IADLs: | Instrumental Activities of Daily Living scale |
| RQ: | research question |
| VLE: | virtual learning environment |
| VR: | virtual reality |
Design Considerations for an Exergame-Based Training Intervention for Older Adults With Mild Neurocognitive Disorder: Qualitative Study Including Focus Groups With Experts and Health Care Professionals and Individual Semistructured In-depth Patient Interviews

Patrick Manser¹, MSc; Manuela Adcock-Omlin¹, DSc; Eling D de Bruin¹²,³, PhD

¹Motor Control and Learning Group – Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
²Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
³Department of Health, OST – Eastern Swiss University of Applied Sciences, St.Gallen, Switzerland

Corresponding Author:
Patrick Manser, MSc
Motor Control and Learning Group – Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich
Leopold-Ruzicka-Weg 4, CH-8093 Zurich, Switzerland
Zurich, CH-8093
Switzerland
Phone: 41 79 519 96 46
Email: patrick.manser@hest.ethz.ch

Abstract

Background: Exergames have attracted growing interest in the prevention and treatment of neurocognitive disorders. The most effective exergame and training components (ie, exercise and training variables such as frequency, intensity, duration, or volume of training and type and content of specific exergame scenarios) however remain to be established for older adults with mild neurocognitive disorders (mNCDs). Regarding the design and development of novel exergame-based training concepts, it seems of crucial importance to explicitly include the intended users’ perspective by adopting an interactive and participatory design that includes end users throughout different iterative cycles of development.

Objective: This study aimed to determine the capabilities, treatment preferences, and motivators for the training of older adults with mNCD and the perspectives of individuals on training goals and settings and requirements for exergame and training components.

Methods: A qualitative study including expert focus groups and individual semistructured in-depth patient interviews was conducted. Data were transcribed to a written format to perform qualitative content analysis using QCAmap software.

Results: In total, 10 experts and health care professionals (80% females) and 8 older adults with mNCD (38% females; mean age 82.4, SD 6.2 years) were recruited until data saturation was observed.

Conclusions: The psychosocial consequences of patients’ self-perceived cognitive deterioration might be more burdensome than the cognitive changes themselves. Older adults with mNCD prefer integrative forms of training (such as exergaming) and are primarily motivated by enjoyment or fun in exercising and the effectiveness of the training. Putting the synthesized perspectives of training goals, settings, and requirements for exergames and training components into context, our considerations point to opportunities for improvement in research and rehabilitation, either by adapting existing exergames to patients with mNCDs or by developing novel exergames and exergame-based training concepts specifically tailored to meet patient requirements and needs.

(JMIR Serious Games 2023;11:e37616) doi:10.2196/37616

KEYWORDS
cognition; exercise; exergame; design; development; neurosciences; technology; training
**Introduction**

**Background**

The normal aging process is associated with a decline in physical and cognitive abilities [1,2]. When the cognitive decline exceeds the normal age-related cognitive decline but is not severe enough to interfere with independence in activities of daily living, it can be classified as “mild cognitive impairment” (MCI), representing an intermediate stage of cognitive impairment between the normal aging process and dementia [3-9]. The condition MCI has evolved over the last decades [5] and has recently been incorporated in the latest Diagnostic and Statistical Manual of Mental Disorders 5th Edition (DSM-5) and the International Classification of Diseases 11th Revision, referred to as mild neurocognitive disorder (mNCD) [7,9-11]. The prevalence of mNCD increases with age, while the incidence of mNCD and the progression to dementia is expected to rise, largely because of the globally growing life expectancies and sedentary lifestyles [3,5,12-15]. As currently no effective pharmacological interventions for patients with mNCD exist [16], alternative options to prevent and treat neurocognitive disorders are needed. Targeting modifiable risk factors in midlife may hold promise for mitigating or even preventing neurocognitive disorders in later life [17-21]. The modifiable risk factors for mNCD include the presence of vascular risk factors (ie, metabolic syndrome, hypertension, hyperlipidemia, coronary heart disease, diabetes mellitus, or stroke) [22-24] or a physically or cognitively sedentary lifestyle [25,26]. Consequently, changes in lifestyle that increase physical and cognitive activity and reduce vascular risk factors are protective against cognitive decline [27-35].

Exergames have gained growing interest to prevent and treat neurocognitive disorders [36-38]. “Exergaming is defined as technology-driven physical activities, such as video game play, that requires participants to be physically active or exercise in order to play the game” [39]. One of the major advantages of exergame-based training is that it is widely accepted by individuals with neurocognitive disorders. In addition, it increases training adherence and engagement by facilitating training motivation and satisfaction [40], which in turn may have a positive effect on the effectiveness of improving cognitive functioning [41]. Furthermore, exergames can be used as a form of simultaneous cognitive-motor training with incorporated cognitive task demands [42]. Meta-analytic evidence suggests that simultaneous motor-cognitive training is the most effective type of training for improving cognition in healthy older adults (HOA) [43,44] and older adults with mNCD [44-46]. For exergames specifically, a recent systematic review synthesized evidence from low risk of bias studies showing that there were consistent positive effects favoring exergaming in people with mNCD and dementia [40]. Nonetheless, it is currently difficult to draw reliable conclusions about the effectiveness of exergaming in preventing and treating neurocognitive disorders because of the substantial variations in the exergame-based training used. Therefore, further investigations are needed for the establishment of effective exergame and training components (ie, exercise and training variables such as the frequency, intensity, duration, or volume of training and the type and content of specific exergame scenarios) for cognitive functioning that can be applied with confidence in evidence-based exergame interventions [36].

Regarding the design and development of novel exergames, it seems crucial to explicitly include the intended users’ perspectives [47]. Taking the characteristics, needs, and experiences into account should ensure adequate use and therefore the effectiveness of the solution. Baquero et al [48] pointed out that an end user–centered methodological design is most often adopted in the development of computer-based training programs for cognitive rehabilitation of older adults with neurocognitive disorders (NCDs). In an ideal case, this process fulfills the international standards for the development of programs including (1) understanding and specifying the context of use (type, characteristics and tasks of users, and physical or social environment), (2) specifying the user requirements, (3) producing design solutions, and (4) evaluating the design [48,49]. So far, only half of the studies reporting computer-based interventions took the standard “specification of user requirements” into account [48]. This has led to the recommendation that future studies in this field should use an interactive and participatory design that explicitly includes end users throughout different iterative cycles of development [48]. In short, it is important to systematically and thoroughly investigate the specific user requirements and preferences for an exergame-based training concept before it is designed and developed.

**Objectives**

This study aimed to determine the capabilities, treatment preferences, and motivators for the training of older adults with mNCD and the perspectives of individuals on training goals and settings and requirements for exergame and training components.

**Methods**

**Overview**

This study is part of the national project “Brain-IT,” which began in August 2020 in Switzerland. The aims of the overall project are (1) to determine the most suitable components for exergame-based training in older adults with mNCD; (2) to explore novel strategies for a real-time adaptive exergame system to individually tailor exergame demands according to users’ physical or cognitive capabilities; (3) to incorporate the acquired knowledge into an exergame-based training concept with the aim of halting or reducing cognitive decline and improving quality of life; and (4) to evaluate the effectiveness of the resulting training intervention in older adults with mNCD. The project is guided by a theoretical framework that provides specific guidance in the design, development, and evaluation of exergames for older adults, the “Multidisciplinary Iterative Design of Exergames (MIDE): A Framework for Supporting the Design, Development, and Evaluation of Exergames for Health” [50], which provides specific guidance in the design, development, and evaluation of exergames for older adults. This study is part of the first phase of the project, with the aim to specify a “set of design requirements that includes design considerations, accessibility recommendations, user modeling.
elements, and technological reflections to be followed in the design and development phase” [47,50], and it was combined with an extensive literature review and reflections on technology. For the project, the exergame device “Senso (Flex)” (Dividat AG) was preselected on the basis of (1) our previous research, (2) because this device has already been shown to be feasible and well-accepted in geriatric patients [51] and patients with major neurocognitive disorder [52], and (3) because it is already widely used (and therefore available more widely and for longer term by end users and health care institutions) for motor-cognitive training within geriatric populations, physiotherapies, or rehabilitation clinics in Switzerland. On this basis, this qualitative study was designed to achieve the defined objectives in general; in addition, it also aimed to collect evidence about the previous experiences of experts or health care professionals with different exergame systems (including the “Senso (Flex)”).

In this way, the project team wanted to collect evidence to make an informed decision whether the specific exergame device was suitable for the project, what possible modifications might be needed to optimize the exergame experience for patients with mNCD, and whether and what alternative exergame devices are suggested by the experts (see subsections of “(T2) Treatment Experience and Preferences—Previous Experiences with Exergames (‘Senso’ specifically)” and “(T5) Exergame and Training Components—Exergame System and Content” in the focus group discussions). Other than parts of these 2 sections that include device-specific findings, none of the remaining sections in this manuscript are device specific.

**Study Design**

A qualitative study was conducted between November 2020 and January 2021, including expert focus groups and patient interviews; both were organized as semi-structured, in-depth interviews. Semistructured, in-depth interviews are the most widely used interviewing format for qualitative research and are generally organized around a set of predetermined open-ended questions, with additional questions and discussion points emerging from the dialogue [53]. The study was planned and reported in accordance with the “consolidated criteria for reporting qualitative research (COREQ)” [54].

The MIDE Framework [50] guided our approach. On the basis of this framework, we integrated multiple stakeholders into the design and development process including exergaming researchers, clinical experts with different backgrounds, a company representing the exergaming industry, and the end users.

**Ethics Approval**

All the study procedures were performed in accordance with the Declaration of Helsinki. The study protocol (not registered) was approved by the ETH Zürich Ethics Commission (EK 2020-N-154). All interested individuals were fully informed of the study procedures. The expected benefits and risks of the study were explained by the study investigator, who was also able to answer open questions and clarify individuals’ uncertainties. It was further verified that withdrawal was permitted at any time during the study without providing any reason. After sufficient time, suitable individuals willing to participate in the study provided written informed consent and were included in the study. No compensation was provided to the participants.

**Participants**

**Experts**

Recruitment aimed at including experts and health care professionals experienced with exergame training of older adults with mNCD, preferably (but not necessarily) with the exergame training system “Senso (Flex)” or similar. For this purpose, Dividat AG was asked to provide a contact list of 10 to 15 external experts and health care professionals with a variety in age, sex, educational level, and experience in therapy of older adults with mNCD, who are not employed by Dividat AG or had received any funds from Dividat AG for their work. All recommended experts and health care professionals were contacted via email between November and December 2020. By applying broad inclusion criteria, a rich spectrum of experts and health care professionals were considered in the study, which in turn will foster the usability of the resulting program in clinical practice. The specific eligibility criteria comprised the following aspects: (1) experts or health care professionals (e.g., physical therapists, movement therapists, neuropsychologists, or researchers experienced with exergames) experienced with exergame training or with older adults with mNCD; (2) German or English speaking; and (3) age ≥18 years. There were no specific exclusion criteria.

**Older Adults With mNCD**

Older adults with mNCD were consecutively recruited between November 2020 and January 2021 in collaboration with health care institutions and (memory) clinics in the larger area of Zürich. Leaflets and study information sheets containing researchers’ contact details were handed out to suitable patients by their therapists. Suitable patients were identified from medical records and patient registries of memory clinics or from diagnostics that had just been performed. Interested patients were contacted by the research team by telephone or email to clarify or obtain further information about the study procedures and to register interest in participating in the study. Subsequently, all patients were fully informed about the study procedures in a face-to-face meeting at the patient’s homes. In addition, patients of interest were screened for eligibility. The eligibility criteria are presented in Textbox 1.
Procedures and Data Collection

Expert Focus Groups

The expert focus groups were moderated by the first author (PM) into groups of up to 5 experts. The moderator was a male doctoral student with a master’s degree in Health Sciences and Technology (ETH Zürich, Switzerland), who was trained for qualitative research. Owing to the COVID-19 pandemic, all focus group sessions were held as web-based meetings in the form of Zoom sessions (Zoom Video Communications), took approximately 60 to 90 minutes to complete, and were audio recorded. Each session started with a short presentation of the background and overall aim of the project. Subsequently, the aim of this study was presented before starting the focus group discussions. The focus group discussions were organized as semistructured, in-depth interviews along an interview guide (Multimedia Appendix 1) [53]. The interview guide was not pilot-tested, as it was developed by the first author (PM) in collaboration with the second author (MA), an experienced clinical neuropsychologist. After questioning the patients’ capabilities as well as their previous treatment or training experience and preferences, the interview continued with questions about motivators for training and the preferred components of exergame-based training. Open-ended questions were asked to enable an open conversation [53]. To focus the moderator’s attention on patients’ verbal and nonverbal communication, no notes were taken during the interviews [57]. Finally, the interviewer was prepared to tailor the interview questions and communication style to the patients’ capabilities, and in case of higher levels of impairment, to adopt strategies suggested to optimize communication with patients with NCDs [58,59].

Sample Size

The intended sample size was set at approximately 5 to 10 experts for the focus group sessions and 5 to 10 older adults with mNCD for the patient interviews; however, study participants were consecutively included until data saturation was reached [60].

Data Analysis

First, all audio files were transcribed in written format in Microsoft Word in pseudonymized form. The transcripts were not returned to the participants for corrections or comments. To explore the perspectives of patients and experts or health care professionals, a qualitative content analysis was performed besides the interviewer, the patient, and, if requested, a care professional or partner as personal support for the patient. We did not set a time limit for the interviews but gave all participants enough time to share their views on the topics discussed. On average, each session took approximately 20 to 30 minutes to complete and was audio recorded. Before starting the interview, the background and overall aim of the project as well as the aim of this study were explained to each patient. The interviews were organized as semistructured, in-depth interviews along an interview guide (Multimedia Appendix 1) [53]. The interview guide was not pilot-tested, as it was developed by the first author (PM) in collaboration with the second author (MA), an experienced clinical neuropsychologist. After questioning the patients’ capabilities as well as their previous treatment or training experience and preferences, the interview continued with questions about motivators for training and the preferred components of exergame-based training. Open-ended questions were asked to enable an open conversation [53]. To focus the moderator’s attention on patients’ verbal and nonverbal communication, no notes were taken during the interviews [57]. Finally, the interviewer was prepared to tailor the interview questions and communication style to the patients’ capabilities, and in case of higher levels of impairment, to adopt strategies suggested to optimize communication with patients with NCDs [58,59].

Sample Size

The intended sample size was set at approximately 5 to 10 experts for the focus group sessions and 5 to 10 older adults with mNCD for the patient interviews; however, study participants were consecutively included until data saturation was reached [60].

Data Analysis

First, all audio files were transcribed in written format in Microsoft Word in pseudonymized form. The transcripts were not returned to the participants for corrections or comments. To explore the perspectives of patients and experts or health care professionals, a qualitative content analysis was performed

Textbox 1. Description of all eligibility criteria.

- **Inclusion criteria**
  - Participants fulfilling all the following inclusion criteria were eligible:
    - (1= mild neurocognitive disorder [mNCD]) clinical diagnosis of “mNCD” according to International Classification of Diseases 11th Revision [7] or Diagnostic and Statistical Manual of Mental Disorders 5th Edition [9]) OR (2= MCI). Patients “screened for MCI” according to the following criteria: (1) informant (ie, health care professional)-based suspicion of mild cognitive impairment confirmed by (2) an objective screening of mild cognitive impairment based on the German version of the Quick Mild Cognitive Impairment Screen [55] with (b1) a recommended cutoff score for cognitive impairment (mild cognitive impairment or dementia) of <62/100 [56], while (b2) not falling below the cutoff score for dementia (ie, <45/100 [56])
  - German speaking

- **Exclusion criteria**
  - The presence of any of the following criteria led to exclusion:
    - Presence of additional, clinically relevant (ie, acute or symptomatic) neurological disorders (ie, epilepsy, stroke, multiple sclerosis, Parkinson disease, brain tumors, or traumatic disorders of the nervous system)
    - Presence of any other unstable or uncontrolled diseases (eg, uncontrolled high blood pressure, progressing or terminal cancer, etc)

Patient Interviews With Older Adults With mNCD

The patient interviews were conducted individually with each patient by the first author (PM) and either took place at ETH Zürich (Institute of Human Movement Sciences and Sport, Leopold-Ruzicka-Weg 4, 8093 Zürich) or at the patients’ homes, depending on the patients’ preferences. The interview sessions were held face-to-face in a quiet room with no one present.

https://games.jmir.org/2023/1/e37616

JMIR Serious Games 2023 | vol. 11 | e37616 | p.277
according to Mayring et al [61,62] using QCAmap software [62-64]. The first step in the analysis involved repeated readings of the transcripts and listening to the original audio files to gain a better understanding of the conversation content. Second, the type of analysis (ie, category assignment procedure) was predefined for each of the research questions (ie, key questions of the interview guide). In case of an inductive category assignment procedure, a selection criterion and level of abstraction were defined for each of the research questions. For deductive category assignments, each research question was operationalized into categories, and a corresponding coding guideline (ie, category label, category definition, anchor example, and coding rules) was formulated. On the basis of this, all transcripts were coded line-by-line (ie, including a revision of the category system after a pilot loop). Subsequently, each resulting list of categories was grouped into main categories, and inter- and intra-agreement checks were performed. Finally, the results of each key question were analyzed along the structure (including predetermined themes and topics) of the interview guide that was created according to the guidelines of the MIDE Framework [50]. Thus, the results were structured and analyzed in 2 main themes and 5 topics. First, the section “user modeling” that included 3 topics: (T1) capabilities of older adults with mNCD, (T2) treatment experiences and preferences, and (T3) motivators for training. Second, “therapeutic needs,” including (T4) training goals and outcomes and (T5) exergame and training components. Within the topic “(T1) capabilities of older adults with mNCD,” the described cognitive capabilities and difficulties were classified into the key neurocognitive domains (as defined by Sachdev et al [10]) in line with DSM-V [9] on agreement between the first (PM) and second author (MA; an experienced neuropsychologist). Within the topic “(T3) motivators for training,” the motivators for training were coded and analyzed against the background of the “Self-determination Theory” [65]. The Self-determination Theory [65] accounts for the quality of different levels of motivational regulation in physical activity settings. It is considered useful to gain a better understanding and promote training motivation, enjoyment, and adherence and has demonstrated considerable efficacy in explaining exercise motivation and behavior [66-70]. Data from the qualitative content analysis were combined with quantitative data (ie, frequency of various statements [f] and in the case of patient interviews, the proportion of patients making a statement [in %]) [60]. The coding and data analysis process was cross-checked to enhance the credibility of the analytic procedure [60].

Results

Participants

In total, 11 external experts and health care professionals were contacted by the first author (PM). All experts responded and were interested in participating. Two experts could not participate in the focus group sessions because of time constraints. According to the “integrative” contribution of the MIDE Framework, “perspectives of various stakeholders (e.g., industry partners, data analysts, health care professionals) are considered in the process of designing and developing exergames” [50]. In accordance with this, the founder of Dividat AG was involved in one of the focus group discussions as an industry representative. In total, 10 experts and health care professionals (80% females) participated in 1 of the 5 focus group sessions until data saturation was observed and further recruitment was terminated. The focus group sessions were conducted in groups of between 1 (k=3) and 3 (k=1) experts (median 1.5) and the moderator (PM). The professional backgrounds of the experts and health care professionals included exergaming researchers (n=4), physical and occupational therapists (n=2), neuropsychologists (n=2), project manager therapy (n=1), and founder of an exergaming company (n=1).

For the patient interviews, 8 patients (38% females; mean age 82.4, SD 6.2 years; mean level of cognitive functioning, measured by the German Version of the Quick Mild Cognitive Impairment Screen [55], 56.0, SD 8.2) were invited and interviewed until data saturation was observed and further recruitment was terminated. None of the patients refused to participate or dropped out of the study after providing their written informed consent. The demographic characteristics of the patients are summarized in Table 1.
### Table 1. Demographic characteristics of the study population.

<table>
<thead>
<tr>
<th>Demographic characteristic</th>
<th>Total sample (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>82.4 (6.2)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>23.1 (2.4)</td>
</tr>
<tr>
<td>Physical activity (min/week), mean (SD)</td>
<td>298.8 (227.0)</td>
</tr>
<tr>
<td>Qmci⁴ [55] total score (points), mean (SD)</td>
<td>56.0 (8.2)</td>
</tr>
</tbody>
</table>

Clinical subtype, n (%)

- mNCD⁵ due to Alzheimer disease: 6 (75)
- Mild frontotemporal NCD: 0 (0)
- mNCD with Lewy bodies: 2 (25)
- Mild vascular NCD: 2 (25)

---

⁴Qmci: Quick Mild Cognitive Impairment Screen.
⁵mNCD: mild neurocognitive disorder.

### Qualitative Content Analysis Results

#### T1: Capabilities

The experts described a large variety of impairments observed in older adults with mNCD. The most frequently described impairments referred to cognitive functioning (f=43), including impairments in the following neurocognitive domains: executive function (f=23), complex attention (f=11), learning and memory (f=7), visuospatial skills (f=2), language (f=1), and social cognition (f=1). These cognitive changes were also described as affecting psychosocial factors (f=22), mainly by causing psychological distress (f=9) and feelings of insecurity (f=2), leading patients to try to hide their impairments from others (f=2). In addition, an increased fall risk (f=9) and reduced physical resilience (f=7) were observed. Although experiencing difficulties in activities of daily living (ADLs; f=1), patients were described as maintaining their functional independence in ADL (f=2).

In line with the experts’ viewpoint, cognitive deterioration (f=22, n=7, 88%) was frequently described by the patients, mainly affecting learning and memory (f=11, n=4, 50% of patients), executive function (f=6, n=4, 50% of patients), and complex attention (f=5, n=2, 25% of patients), whereas only minor restrictions in physical capabilities and mobility were mentioned (ie, impaired balance, [f=2, n=2, 25% of patients], reduced gait speed [f=1, n=1, 13% of patients], increased fall risk [f=9, n=5, 63% of patients], fatigue [f=6, n=3, 38% of patients], and joint pain [f=2, n=2, 25% of patients]). ADLs remained preserved in all patients, but the need for coping strategies was mentioned by 4 patients (50%) to be able to preserve ADLs. From the patient’s perspective, the consequences of their self-perceived subjective cognitive decline (f=8, n=6, 75% of patients) with regard to psychosocial factors were most frequently reported (f=36, n=8, 100% of patients), mainly involving psychological distress (f=13, n=2, 25%), feelings of insecurity (f=6, n=3, 38% of patients), depressive symptoms (f=2, n=2, 25% of patients), or fear of repeated falls (f=3, n=1, 13% of patients):

> A really tedious thing is that you often can’t keep up. For example, in discussions or conversations. [...] You often think about what the other(s) have just said and in the meantime he or she has already continued. That’s why you often just don’t say anything. Of course, most people like it when you don’t say anything (*laughs*). So, these people don’t get upset about it. But I am. [P-01]

> I used to go running a lot. I don’t do that anymore. But swimming is still fine. In the worst case, I become a drowned corpse, but at least I can’t fall while swimming. [P-02]

> I can actually do everything; I just have to be careful because of my dizziness and weakness so that I don’t fall. I also have problems with short-term memory. I have to try to remember everything somehow, but I still forget a lot of things. [P-04]

#### T2: Treatment Experience and Preferences

### Previous Treatment and Training Experiences

To counteract cognitive decline and preserve physical capabilities, mobility, and ADLs, patients have already been on medical training therapy (MTT; f=3, n=3, 38% of patients), have already been on physical therapy (PT; f=2, n=2, 25% of patients), have performed a specific group-based (ie, f=1, n=1, 13% of patients) or individual (f=1, n=1, 13% of patients) cognitive training or meditation (f=1, n=1, 13% of patients), or have reported to have no experience in any specific therapy or training (f=1, n=1, 13% of patients).

From the patient’s viewpoint, MTT and PT were perceived as useful (f=3, n=3, 38% of patients), but patients reported that they would have to do it more consistently to profit from it (f=2, n=2, 25% of patients). Computerized cognitive training (CCT) was also perceived as useful (f=1, n=1, 13% of patients) and reported to be challenging, fun, and enjoyable (f=2, n=2, 25% of patients). Nonetheless, patients reported being insecure about the effectiveness of CCT (f=2, n=2, 25% of patients):
According to the experience of experts and health care professionals, only cognitive forms of training or physical exercises were often experienced as boring over time by older adults with mNCD (f=2) and required guidance by a therapist (f=1). More integrative forms of training, including gamified tasks close to everyday life, multimodal animation, and acoustic feedback, were reported to be preferred by patients (f=4):

**Previous Experiences With Novel Technologies**

Although being described as skeptical about the use of technological devices, experts perceived older adults with mNCD as ready to use technological devices such as heart rate monitors during training (f=9), if its usability is ensured:

**Previous Experiences With Exergames [“Senso (Flex)” Specifically]**

None of the interviewed patients reported any previous experience with exergames in general or with the exergaming system “Senso (Flex)” specifically. Nonetheless, after a short introduction to the system, all patients stated that they would be willing to try it.

On the basis of the previous experiences of the experts and health care professionals, the interaction with the “Senso,” its overall usability, and the design of the exergames have been described as good (f=5). Regarding hardware components, minor usability problems have been reported. Patients were observed to unintentionally walk off the middle plate without noticing the feedback on the screen (f=1), constantly change their focus between the game tasks on the screen and the stepping plate to anticipate and plan their movements (f=1), or make too small steps to tap on one of the outer stepping plates (f=1). In addition, the patients needed time to familiarize themselves with the sensitivity of the stepping plate (f=2):

**The tasks on the “Senso” are very well designed.** [E-08: project manager therapy]

**Most of the time, the patients look down at that very square in the middle of the “Senso”.** [E-09: exergaming researcher]

Additional usability issues were reported to be linked to the capabilities of older adults with mNCD. First, it has been described that patients are often cognitively overloaded when trying out new games (f=1), by the occurrence of an unexpected situation or technical errors (f=2), or by the cognitive task demands required to interact with the exergame system in general (f=1), which may limit training duration owing to attentional exhaustion (f=2):

**With new games, patients are often overwhelmed in general, because they don’t know what to expect.**
They often need time to find their way around. [E-07: physical and occupational therapist]

[...] Patients are completely overwhelmed as soon as something unexpected or a technical problem occurs. [E-03: neuropsychologist]

In contrast, the physical capabilities were reported to not directly affect the usability of the system (f=4), although some patients experienced difficulties with backward steps (f=2), and many patients made use of the handrail to reduce the physical strain (f=6). In some cases, physical limitations (eg, fatigue and joint pain) resulting from static loading have been reported to limit the training duration (f=4):

Patients often have problems with backward steps. [and] Patients hold on to the handrail far too often. [...] it is often the case that people hold on because it is simply ‘a bit more comfortable. [E-10: founder of an exergaming company]

Often it is already difficult and tiring for patients to stand for a longer period of time. It is often easier for them to walk. [and] However, it should be noted that this form of fatigue is not necessarily comparable to fatigue caused by physical training. Fatigue does not necessarily come from physical exertion. It is possible that this type of fatigue is caused by the static load and the resulting joint pain. [E-06: physical and occupational therapist]

When considering the specific games of the exergaming device “Senso” (video illustrations and explanations of all currently available games can be found at [71]), the simple and clear design structures of the games (f=4) and the intuitive tasks were reported to be highly appreciated by patients and promote good comprehensibility, which was reported for the games “Simple” (f=3), “Birds” (f=3). Nonetheless, there are also games that were reported to cause problems of understanding, in particular the games “Simon” (f=3), “Tetris” (f=3), “Habitats” (f=4), “Targets” (f=1), and “Snake” (f=2). These problems may be related to the game instructions (f=9):

[...][Many people are very happy with simple design structures. This should be maintained at all costs when designing new games for MCI patients. However, [...] some kind of adjustment of the game instructions is definitely needed. [E-10: founder of an exergaming company]

For patients, a game does not stand out by its great graphics, but by the game tasks as such. [E-08: project manager therapy]

[About problems of understanding the games] I think the reasons were that they didn’t really understand the instructions well. [E-09: exergaming researcher]

However, it could also be related to the task demands of the games. It was reported that the patients need some time to familiarize themselves with the game to fully understand it (f=1). According to the experts’ experiences, this works well with the games “Simple” (f=4), “Birds” (f=1), “Flexi” (f=1), and in some cases “Habitats” (f=1). At the same time, games such as “Flexi” (f=1), “Habitats” (f=6), “Hexagon” (f=3), “Simon” (f=6), “Ski” (f=4), “Targets” (f=12), and “Tetris” (f=4) were frequently reported to start at an already (too) challenging level for older adults with mNCD and progress too fast while there is a limited range of games or adaptability of task demands at the lower end of difficulty levels (f=9). This was mentioned to be mainly apparent for the cognitive task demands (eg, game speed and task complexity), whereas physical exertion intensity is often (too) low and could be increased (f=4):

For MCI-patients, some games are predestined to be used with them, such as “Simple,” “Flexi,” “Birds” and perhaps also “Habitats.” These games don’t put so much time pressure and the feeling of having missed something on patients. [E-08: project manager therapy]

[...] the increase in the challenge profile from the easiest games (“Simple” and “Birds”) to the next more difficult game is too steep for MCI-patients. For example, the game “Targets” is too fast for many patients. The game “Habitats” contains too many stimuli at once, so that the patients no longer know what they have to pay attention to. [E-07: physical and occupational therapist]

One problem with the “Senso,” in general, is that the physical intensity might well be higher. [E-05: neuropsychologist]

Overwhelming task demands may cause frustration or refusal of games (f=6), although the feedback mechanisms to indicate errors work subtle (f=4). In contrast, games that are perceived as being too easy lead to boredom (f=2):

For example, the games “Targets,” “Ski” or “Hexagon” are very confronting, and patients recognized quite quick: “Okay, I can’t do it,” and that frustrates patients. [...] Usually, these patients stop in the middle and say something like: “Ah, I don’t need that kind of shit.” Most of the time, they stop the training session immediately and don’t want to continue anymore. [E-08: project manager therapy]

My observation was that the negative feedback currently used does not demotivate the patients at all. It is also clear to the patients that they need to know when they are making mistakes and whether they are completing the tasks correctly. [E-04: exergaming researcher]

Some of the negative feedback is so subtle that it is not even noticed. [E-05: neuropsychologist]

**T3: Motivators for Training**

The experts described numerous motivators for training older adults with mNCD. The most frequently described motivators can be classified as intrinsically regulated motivators (f=44), which are directly related to exergames. Excitement, enjoyment,
or fun is perceived as a central motivator for performing exergames (f=4). This was reported to be maintained by the captivating character of exergames (f=1) and multimodal animation (f=1), which is supported by specific game components (eg, game tasks or designs close to everyday life [f=6] or with personal relations or memories [f=1] including music or sound effects [f=8], animals or plants [f=4], landscapes [n=1], or colors [f=1]). In addition, patients were described as intrinsically motivated by gamification (f=6), the feeling of being optimally challenged (n=3), or simply by the variation of training (f=6):

For patients, the focus is primarily on having fun with the games. For example, they [... liked watching birds and listening to birdsong and felt very motivated by the personal connection. Through these personal memories [...] a whole other level of motivation emerged. [E-08: project manager therapy]

I think that those people who enjoy playing games are generally captured by the playful and competitive nature of the games. Furthermore, training with exergames is something completely different compared to classical therapy. Patients appreciate this change from the “dry” standard therapy. [E-06: physical and occupational therapist]

However, when task demands become too high (f=6) or too low (f=2), patients have been observed to promptly lose their willingness to perform the exergames, as already reported. External motivators such as social support (eg, by therapists or caregivers) or group dynamics have also been reported to improve motivation to train (f=12). Feeling concerned about cognitive deterioration or being confronted by performance classifications can either motivate or induce negative feelings (f=7). Finally, some patients were also reported to be motivated by the effectiveness of exergames (f=2) or performance improvements (f=2):

I consider this social support to be very central. [...] If a relative joins in for motivation or support it can be very valuable. [E-04: exergaming researcher]

I think there are always patients who don’t want to know how well they are performing. Forcing performance feedback on such people can of course be motivating, but it could also be negative and confirm their limitations. [E-07: physical and occupational therapist]

From the patients’ viewpoint, all patients reported that they could primarily be motivated to train regularly by the effectiveness of the training, helping them achieve their individual success (f=13, n=8, 100%). Alternately, patients reported being motivated by their relatives or partners (f=2, n=1, 13%) and enjoyment of exercising (f=1, n=1, 13%). Having to travel to a training facility was reported to have a negative effect on training motivation and adherence (f=4, n=1, 13%):

It would be nice if I could go for a walk in the forest again without falling down. I used to do this four times a week for 75 minutes. It motivates me to train so that I can do this again in the future. [P-02]

It would motivate me if I could improve my abilities (balance) again. [...] I would like to stay independent and modern, not to be called an old lady. [P-03]

The success. I no longer need to be motivated. If I set my mind to it, I do it! [P-08]

T4: Training Goals and Outcomes

Regarding the training goals, cognitive functioning (f=19) should be targeted in the training intervention in the experts’ viewpoint while also addressing ADLs and mobility (f=3), addressing physical capabilities (f=3), and accounting for psychosocial factors (f=2), such as feelings of insecurity. However, the weighting of the training focus differs significantly between experts in different fields:

[...] higher order processes (i.e. divided attention or the ability to plan) are affected in most patients. Therefore, it is important to focus on these higher order cognitive functions. [E-05: neuropsychologist]

I think that the coupling of brain functions with physical functions is central. At the same time [...] it is important to focus on what is impaired. If the frontal lobe is impaired, it is certainly important to train executive functions, attention and inhibition. [E-10: founder of an exergaming company]

Primarily physical activation, especially that people get moving and walk. But also, to train the intuitive way of taking steps. [...] The cognitive aspects of the training have always played a subordinate role for me, but they were usually not decisive for the success of the therapy, as this was often trained differently, and I am not an expert in this. [E-06: physical and occupational therapist]

When asking experts about the training goals of patients, ADLs and mobility (f=5) were the most frequently stated in addition to cognition (f=3) and physical functioning (f=2). In addition, psychosocial factors (f=2) have been reported to include socializing or having fun:

I had patients who wanted to continue training because the training made them more confident in their gait. They felt better balance after the training. [E-06: physical and occupational therapist]

The patients also see the cognitive aspects of the training, of course. [...] We often explain to the patients that falls prevention has a cognitive and physical aspect and that these aspects interact. Therefore, the patients mainly go to the training with the aim of improving their gait. [E-07: physical and occupational therapist]

Some people really know what’s going on and they know that they have a disease and that they can prevent or slow down the progression by doing physical activity and exergames. But then others don’t really know that they have cognitive deterioration and they’re just playing a game and having fun without specific training goals. [E-09: exergaming researcher]
This is consistent with patients’ viewpoint who most frequently reported improving gait (f=6, n=5, 50%), memory (f=3, n=3, 38%), and balance (f=2, n=2, 25%) as their primary goal to increase their quality of life. In addition, patients reported being more active (f=1, n=1, 13%), increased functional abilities (ie, cooking; f=1, n=1, 13%), increased strength (f=1, n=1, 13%), or remaining independent in ADLs (f=1, n=1, 13%) as training goals:

- It is mainly the memory. It is memory because it affects a lot of other things. [P-01]
- It would be wonderful, if I could go for a walk in the forest again without falling down. [P-02]
- I really want to remain independent. I definitely don’t want to become dependent on others. [P-05]
- That I can keep things better in my head. That has diminished. That would be nice! [P-06]
- I want to have more strength again to increase stability and be able to walk longer. [P-08]

**T5: Exergame and Training Components**

**Location**

Regarding training location, the experts reported that the patients would either prefer individual training at home (f=3) or in a mixed setting, including training at home combined with training at a clinic (f=4). None of the experts stated that patients would prefer exercising at a clinic or training facility in general, as this is often associated with excessive time expenditure. Training at home was reported to be beneficial, because it represents a known environment that makes patients feel more secure. However, the experts also stated that patients may not be capable of performing exercises or exergames independently and therefore need guidance throughout each training session (f=4) or at least partially (f=9); for example, when starting up the system or in case of technical problems:

> The advantage of training at home is that “it’s a known environment and they feel safer at home and also don’t have to travel.” However, “I would suggest that the help of a guiding therapist with experience will be necessary.” [E-09: exergaming researcher]

> In a previous investigation [...] patients’ feedback was that 70% could imagine doing the training from home. [...] For MCI-Patients specifically, relatives may be involved. But in general, the need for home-based exergame training is there, I would say. [E-08: project manager therapy]

This is also reflected in the outcomes of the question of whether patients would be capable of performing home-based exergame training; the experts mainly reported that patients are certainly capable (f=4) or should be capable of considering some concerns (f=9) to perform such a training program independently at home. The concerns that need to be considered include the improvement of game instructions (f=2), accessibility of a handrail or similar for safety support (f=2), and avoidance of technical problems (f=2) or the integration of a guided familiarization period (f=1) or support of a care professional or partner (f=2):

> I think if the system would really work properly then you could use it at home. However, if you just have some minor technical problems is already like a no-go to use it at home at all. [E-01: exergaming researcher]

> It would certainly be good if the patients could complete an accompanied training for a certain period of time in order to facilitate the transfer to training at home. [E-04: exergaming researcher]

> [...] some kind of adjustment of the instructions is needed [...], especially for this patient group and for independent training in the home-based setting. [...] The instructions have to be adapted in such a way that understanding can be achieved without someone having to stand next to the patients all the time. [E-10: founder of an exergaming company]

Of those patients who responded to the question and had a clear preference regarding the training location, most (6/7, 86%) patients would clearly prefer to train individually at home, because it is less time consuming and more flexible. One patient did not have a clear preference; she simply wanted to perform the exercises where it was easiest for her and preferred group exercises:

> For me, it is important that the training can be done independently at home. If I have to go to the doctor somewhere every time, it’s simply too much work. [P-01]

> Preferably at home, if I can. Then I can also choose the time when I want to exercise. I have lived my whole life with a packed schedule. Now I want to be a little freer and more flexible. [P-03]

**Safety**

The experts reported an increased risk for falls, as patients with mNCD (1) are easily distractable and (2) have difficulties in self-assessment and impaired planning abilities. Therefore, it was recommended to use the handrail in the beginning to minimize the risk of falls (f=3), which was also requested by 1 patient. In the case of a home-based exergaming system—which may not have a handrail—thorough and clear safety instructions are recommended (f=1):

> Especially in the beginning, until the patients have understood what it is all about, it is very important to instruct using the handrail. [E-04: exergaming researcher]

> I definitely need a railing to prevent falls during training. I often fall down if I don’t have anything to hold on to. [P-02]

**Instruction, Familiarization, and Guidance**

As illustrated earlier, certain adaptations are required to enable a more independent use of the exergaming device. First, patients should be familiarized with the exergaming device and the corresponding games considering the following key elements:

1. Start at an easy level (f=7), for example, by using the game “Simple” (f=4), (2) ensure that patients voluntarily try out the device (f=3), (3) ensure that you are not too confronting (f=2), (4) give patients enough time to familiarize with the new task.
It is very important to start very slowly and at a low difficulty level until the patients can better assess their abilities on the “Senso”. [...] Since the game “Simple” waits for a reaction from the individual, it is very suitable to start with. [E-04: exergaming researcher]

We always start with a reaction game so that the patients can learn the coupling of the cognitive-motor functions and learn to interact with the environment. Later on, we focus on specific cognitive functions. [E-10: founder of an exergaming company]

Regarding the instructions, some adjustments are needed to improve comprehensibility. Currently, there is instructional text before starting each game. However, patients with mNCD have been reported to have limited comprehension of instructions. Therefore, adaptations are needed in the instructions of exergames in general and for a home-based exergaming system in particular. The experts mainly suggested to use step-by-step (f=3) instructions based on a combination of visual (ie, written instruction or video demonstration) and verbal instructions (f=4) guided by an experienced therapist (f=1). In case of more severely impaired patients or for home-based exergaming systems, it was suggested that practical demonstrations (f=2), video instructions (f=6) or even interactive “trial run” instructions (f=5) could improve comprehensibility of the games:

The transfer from the written instructions to the understanding of what is to be done in the game is sometimes difficult. [E-06: physical and occupational therapist]

Personally, I would replace the written instructions with a short (few seconds) video sequence showing the most important functions of the games. [E-08: project manager therapy]

I would recommend combining visual and verbal instructions. For example, through a visual presentation with additional step-by-step verbal instructions. Verbally we can “pick up” the patients very well and get a feeling whether the patients have understood the instructions. [E-03: neuropsychologist]

[...], some kind of adjustment of the instructions is needed. [...] It is definitely important to pursue and use these adaptations, especially for this patient group and for independent training in the home-based setting”, because “in the case of more severe impairments, it is often necessary to demonstrate the games step by step by yourself. [...] In other gaming systems there is a short test phase with explanations and trial runs [...]. However, this would have to be offered as an option, since most patients will no longer need it after a few sessions. [E-10: founder of an exergaming company]

Finally, when guiding patients through their training sessions, social support and guidance by a care professional or partner might be beneficial (f=3). However, it was also mentioned that this might be critical because of personal conflicts or patients’ psychological constraints (f=2):

Family members could play an important role in reminding and motivating patients to complete their training. [E-06: physical and occupational therapist]

I don’t think it’s always a good idea to include family members as guidance, because the pressure to perform gets higher for the patients, since they try to hide their impairments from others. A health care professional like a nurse for example or physical therapists would be better than a husband or wife, I think. They already have a lot of fights in the households, because things are not working out as they should. [E-09: exergaming researcher]

From a patient’s perspective, all patients reported that they can imagine training alone, provided they had received thorough instructions and understood their tasks. One patient additionally requested regular support from a care professional or partner:

Yes, I think so. Once I learn that, I’m sure I can do it independently. [P-03]

If I am supported by you or my partner, then I can certainly train partly independently. [P-07]

Exergame System and Content

Previous experiences of older adults with mNCD using the exergaming system “Senso” are illustrated earlier. Building on this, several game-specific adaptations were suggested by the experts (f=9):

More time should be provided between the balls so that the flood of information is reduced (it is often overwhelming when several balls are visible on the screen very quickly). [E-07 (physical and occupational therapist): for the game “Targets”]

In the initial phase, until patients’ have understood all the game tasks [...], the speed must definitely be reduced. [E-06 (physical and occupational therapist): for the game “Habitats”]

There are already enough opportunities to increase the task difficulty. [...] However, it is very important to note that the game difficulty is adjusted downwards so that it is easier to start the training. [E-08: project manager therapy]

In addition to these game-specific adaptations, multiple novel game designs and elements have been suggested and discussed by focus groups to address patients’ needs optimally. In general, it has been recognized that there is a need for new games specifically targeting the neurocognitive functions of learning and memory (f=4) and executive functions (ie, working memory and cognitive inhibition; f=2). Specific game design suggestions were discussed for such a memory or working memory game. Additional suggestions for new game designs and elements include the use of music, addition of visual reminders to guide patients within the games, or adaptations in performance feedback:

With the “Senso”, a certain spectrum of neurocognitive function domains is covered. However,
games for working memory, inhibition or memory are completely missing. In the case of memory, there is currently only one game available specifically targeting the training of short-term memory span. [E-05: neuropsychologist]

I think music would be very motivating for people with MCI or dementia also if it is music from their youth or music they like. It’s also been described in the literature that music has so many good effects on people when they have heard a song that they liked before and they are singing that song. [E-09: exergaming researcher]

In addition, it would be good to include reminders, for example at the edge of the screen, which patients can use for orientation. [...] Additionally, [...] it would certainly be helpful here if the program not only displayed the performance curve, but also provided a reason or explanation. [E-03: neuropsychologist]

As general requirements when designing new games, the experts recommended using simple graphics and ensuring good contrast (f=14), a comfortable relation, and good usability of the exergames (f=4) using easily comprehensible and clearly designed tasks (f=2) with a certain closeness to everyday life (f=7). Multimodal animations, including multisensory feedback (f=7), should additionally be integrated by focusing on positive reinforcement mechanisms (f=2) to motivate the patients during exergaming. In addition, it is important that the main task is in the center of the screen (f=1) and that only elements that are related to the game task are included (f=5). Moreover, too confronting performance feedback (f=1) and unexpected appearance or technical problems (f=2) should be avoided:

It is very important to create a good contrast. [...] It’s generally important for the older population to keep the graphic representation as simple as possible, because for older people, the game is not characterized by great graphics, but by the game task as such. The main importance is that the right level of challenge is offered. [E-08: project manager therapy]

It is much better to present a simple graphic and focus on the aspects that need to be trained. [...] unnecessary graphic gimmicks should be avoided! [E-04: exergaming researcher]

It is important to have a main action that is in the center of the screen and to ensure that the player will have primary task in the center. If you put any secondary tasks into the games, it can be confusing for the patients. [E-02: exergaming researcher]

Spontaneously, I would say that games close to everyday life are more popular. [...] These games were much better received than abstractly structured games (“visual exploration tasks”). [E-03: neuropsychologist]

My experience so far is that games that are designed to be more relevant to everyday life (and simpler) work better. Therefore, new game designs should be based on what patients know from their everyday lives. [E-06: physical and occupational therapist]

Training Components

The recommended exercise frequency ranged from 2 (f=3) to 5 or more (f=4) training sessions per week, largely dependent on training location and motivation. The recommended session durations ranged from a maximum of 15 to 20 minutes (f=3) up to 30 minutes (f=2), with the aim of reaching a moderate exercise volume of approximately 150 minutes per week (f=1). Shorter sessions and a higher training frequency have been reported to be preferable to reach this training volume, mainly owing to attentional exhaustion:

The more the better! I would prefer shorter training sessions, especially because of attentional exhaustion. Here I would recommend a maximum of 30 minutes and at least 5 sessions a week. This is much better than training for 2 hours at a stretch! [E-03: neuropsychologist]

I would recommend a training frequency of 2 – 3x/week. [...] The training duration is difficult to estimate. Some patients are already exhausted after 2 minutes, others can easily train for 20 minutes. [E-10: founder of an exergaming company]

I think that a training frequency of 3x/week is already (too) much. 2x/week should be possible to arrange. 1x/week definitely works. This may be because three appointments, in combination with other activities, may already be too much for patients. If the training could be done at home, the training frequency could certainly be increased up to 4 - 5x/week. In this case, motivation could still be difficult. [E-07: physical and occupational therapist]

I would aim for a training volume of 150 min/week. As far as I know, this is considered moderate for older patients. I would consider 100 min/week as the lower limit. A minimum of 3 x per week for 30 min would also be okay at best. [E-08: project manager therapy]

Exercises requiring a coupling of physical and cognitive functions were described as preferable and should be prescribed domain-specific depending on the patient’s abilities:

I think that the coupling of brain functions with physical functions is central. Whether this is ultimately an attention game, or a training of the executive functions is something I don’t consider central at the beginning. Of course, it also plays a role here which cognitive functions are impaired. [...] If the frontal lobe is impaired, it is certainly important to train executive functions, attention and inhibition. [E-10: founder of an exergaming company]

To maintain the training program in the long term (preferably >12 weeks), motivation is a key factor that can be facilitated by the playful character of the exergames and a variation in the choice of games. Nonetheless, patients seem to prefer a certain routine:

Of course, the training should be maintained over a certain amount of time at a stretch. So not just two
weeks, but ideally longer (more than 12 weeks). Of course, motivation is also a very central point. If the training is varied and has a playful character, this should be feasible. [E-03: neuropsychologist]

Patients are generally routine-oriented, which can also be observed in general. Therefore, it is also important to introduce a new game every now and then. The patients primarily prefer the familiar games and should therefore be challenged to a certain variety. [E-10: founder of an exergaming company]

The physical exercise intensity should be maintained at a light to moderate level, while the focus should be on game complexity that should be challenging but feasible. Game complexity can be varied on multiple levels, for example, (1) stability support (use of handrail with both hands, 1 hand, or no support), (2) stepping direction, (3) game choice and tasks included, (4) game duration, or (5) game speed:

Adding new games. I always start with the game “Simple” and sometimes in the first session I also introduced “Birds” when I think it would be possible. If not, then I will do it the next session. If somebody is really performing well and understanding all the instructions, then I also progress to the game “Targets” and even “Birds”. [E-09: exergaming researcher]

I also often started with just stepping movements forward [...] and included the step direction to the right at a later timepoint. [E-10: founder of an exergaming company]

We have a routine that we usually do the training sessions over 3 weeks and do the first 3 sessions with holding, just to get a feel for the games. After that, we gradually go back to holding on with one arm and without holding on. [E-08: project manager therapy]

From the patients’ viewpoint, a high training frequency (mean preferred training frequency 5.21 times per week; n=7), ranging from 2 times per week (n=1, 13%) to daily sessions (n=4, 50%) with short session durations (mean preferred session duration 23.4, SD 10.3 minutes; n=8), ranging from 10 minutes (n=1, 13%) up to 30 minutes (n=3, 38%) was preferred. Five of 6 (83%) patients who responded to the questions about how long they would prefer to do the training stated that they would prefer to continue the training as long as they profit from it and are able to do it. All patients preferred a training that is individually adapted to apply moderate (4/5, 80% of patients) to high physical (1/5, 20% of patients) intensity and moderate (3/5, 60% of patients) to high (2/5, 40% of patients) cognitive challenges:

If the device was at home, I would do the training every day. [P-01]

I don’t want to make a guarantee now, but I could do a short training session every day for like 20 minutes or so. But I can’t promise that I’ll do 40 minutes every day, because I also want to do other things. Especially when the weather is nice, I like to go outside. And then I also must do the housework, which also takes time. [P-03]

I think about 30 minutes is good. If it goes on too long or is too strict, then I get tired of it. I don’t like that. That would be counterproductive. [P-05]

If I have the device, I could do this training forever. As long as I still have the strength to do it. [P-02]

I would need a bit of a start-up period first. If it’s not effective, I’ll stop again. Additionally, I don’t know how my health will be in the future. But as long as I’m reasonably fit, I’ll definitely want to do it. [P-07]

Individualization

Individualization of the exergame intervention concept should mainly account for two aspects: (1) task type (ie, choice of exergames to individually focus on neurocognitive functioning; f=4) and (2) task demands (ie, adapting the game demands according to the individual capabilities to maintain a challenging but feasible cognitive load; f=5). In addition, it was recommended to change between games with different task demands to enable the maintenance of attention over the entire training duration (f=2) and to supervise training exertion (f=3):

It is important to have a system that will adapt the games according to the participant’s performance. [E-02: exergaming researcher]

The physical intensity is often not a problem, and it should primarily be the complexity of the training that is individually adapted so that it is doable and still has a certain physical demand. [E-10: founder of an exergaming company]

[…] We also have to check whether somebody is very fatigued [...]. Sometimes you have to let someone take a rest because they will not always feel when they have to take a rest. [E-09: exergaming researcher]

One should “[...]alternate between games that focus primarily on performance and less on cognitive aspects with more cognitively demanding games.” [E-08: project manager therapy]

Discussion

Principal Findings

The objective of this study was to determine the capabilities, treatment preferences, and motivators for training older adults with mNCD, as well as their perspectives on training goals, settings, and requirements for exergames and training components. This will—together and in line with a synthesis of the optimal evidence-based informed decisions—serve as basis for user modeling, determination of therapeutic needs, and definition of a set of requirements for the game design and development process of a novel exergame-based training concept. To the best of our knowledge, this is the first study to systematically and thoroughly investigate user requirements and preferences for an exergame-based training concept before it is designed and developed specifically for older adults with mNCD based on these findings.

The results of our qualitative study, which included focus groups with 10 experts or health care professionals and individual semistructured, in-depth interviews with 8 older adults with

Manser et al
mNCD, yielded the following key findings: (1—capabilities) from a patients’ viewpoint, the psychosocial consequences of their self-perceived cognitive deteriorations might be more burdensome than the cognitive changes themselves; (2—treatment preferences) more integrative forms of training (such as exergaming) including gamified tasks close to everyday life, multimodal animation, and acoustic feedback are preferred by patients. None of the interviewed patients reported any previous experience with exergaming, but all patients described the handling of different technologies as feasible despite some challenges and were willing to try out exergaming; (3—motivators for training) from the expert’s viewpoint, the most frequently described motivators to train can be classified as intrinsically regulated motivators such as excitement, enjoyment, or fun in exercising that is maintained by the captivating character of exergames supported by specific game components (eg, game tasks or designs close to everyday life or with personal relations or memories including music or sound effects, animals or plants, landscapes, or colors); the feeling of being optimally challenged; and the variation of training. All patients reported that they could primarily be motivated by the effectiveness of the training, helping them to achieve success on an individual basis; (4—training goals and outcomes) the most important training goals of older adults with mNCD include improvements in ADLs and mobility (gait and balance) and memory, because these outcomes were described as central to improving their quality of life; (5—exergame and training components) the use of home-based exergames as a form of simultaneous-incorporated motor-cognitive training should be prescribed with a domain-specific training focus depending on a patient’s cognitive abilities, a high training frequency (4-5 training sessions per week), short session durations (20-25 minutes), and individual adaption and progression of task type and demands to reach a light to moderate level of physical intensity and a challenging but feasible game complexity. To maintain the training program in the long term (preferably >12 weeks), motivation is a key factor and should be facilitated by the playful character of the exergames, variation in the choice of games, and ensuring that the patients are optimally challenged. To make home-based training interventions feasible, multiple factors that need to be considered were identified. Patient-friendly game instructions are needed, while the accessibility of a handrail or similar for safety support, avoidance of technical problems, and the integration of a guided familiarization period or support from a care person need to be ensured to make home-based exercise training feasible. As general requirements for exergame design, simple graphics with good contrast and easily comprehensible and clearly designed tasks with a certain closeness to everyday life should be used. Multimodal animations, including multisensory feedback that focuses on positive reinforcement mechanisms, should be integrated to motivate patients during exergaming. In addition, it is important that the main task be in the center of the screen and that only elements that are related to the game task are included. Moreover, confronting performance feedback and unexpected appearances or technical problems should be avoided.

### Capabilities of Older Adults With mNCD

A variety of cognitive changes mainly affecting the neurocognitive domains of learning and memory, complex attention, and executive function were discussed by the focus groups and mentioned by the patients, whereas no serious restrictions on physical capabilities, mobility, and ADLs were reported. This is in line with DSM-5 [9]. According to the definition of mNCD, modest (ie, for mNCD, performance typically lies in the 1-2 SD range) deterioration in cognitive functioning can be observed, whereas the capacity for independence in everyday activities is preserved [9]. However, from the patient’s perspective, the most prominent consequences of their disorder were described as affecting psychological factors, mainly by causing psychological distress, feelings of insecurity, and depression. It is well known that depression and anxiety are common in older adults with mNCD [72,73]. In addition, patients with depression have higher rates of conversion to dementia, indicating that depression is an important risk factor for cognitive decline and progression to dementia. This emphasizes the importance of assessing depressive symptoms in older adults with mNCD [72].

### Treatment Experience and Preferences

Most of the interviewed patients had already gained experience with different treatment or training approaches to counteract cognitive decline and preserve physical capabilities, mobility, and ADLs. Although MTT, physiotherapy, and CCT were perceived as useful, the patients reported being insecure about the effectiveness of these approaches or that they would have to (be able to) do it more consistently to profit from it, which was described to be limited by the availability of therapists. More integrative forms of training, including gamified tasks close to everyday life, multimodal animation, and acoustic feedback, were reported to be preferred by patients.

This is in line with the literature, showing that “research involving older adults has found that CCT programs are associated with high satisfaction levels, and that they are also a feasible option for individuals with MCI, with equal or better adherence rates when compared with traditional cognitive training” [74-76]. This is also evident in the use of exergames. Exergame-based training interventions are widely accepted in individuals with mNCD, and exergames increase or enhance participants’ motivation to engage in rehabilitation activities [40]. This is also reflected by the adherence rates to different types of exercises in patients with mild to major NCD. Recent systematic reviews and meta-analyses synthesized mean adherence rates of 70% [77] for physical exercise interventions, whereas the mean adherence rate was higher for exergame-based interventions at 84% [78]. To the best of our knowledge, there is no systematic review that has synthesized adherence rates to CCT. However, Turunen et al [79] investigated adherence to a long-lasting multidomain CCT among a sample of 631 older adults at risk of dementia. It was shown that only 20% of participants completed at least half of their CCT sessions, and only 12% of participants completed all (maximal number of training sessions=144) of their training sessions. In addition, 37% of the participants did not train at all, whereas “previous use of computers, better memory, being married/cohabiting,
and positive study expectations were independently associated with the greater probability of starting the CCT. Previous computer use was the main determinant of the number of CCTs completed after the training was initiated [79]. Therefore, when comparing these findings, it appears that exergame-based interventions have the highest adherence rates among different training programs. This is consistent with findings in HOAs, where adherence to technology-based training programs was higher than that to traditional training programs, independent of study site or level of supervision [80]. This finding may be largely explained by the high level of enjoyment in using technology-based physical exercise programs [80].

Technology-based training systems offer several advantages over traditional training programs that may contribute to a more enjoyable exercise experience. For example, exergames can provide real-time feedback and positive reinforcement while exercising and can monitor performance over time [80]. In addition, exergames enable individual real-time adaptivity of physical and cognitive task demands according to the participants’ performance or physiological response (eg, heart rate and brain activity), which is considered a key advantage of serious video games (such as exergames) [81-83]. In fact, the findings of our study suggest that applying an optimal challenge is central to promote the use of exergames in patients with mNCD in the long term.

When considering the experts’ previous experience in the use of exergames (ie, “Senso”) with patients with mNCD, the interaction with the device, its overall usability, and the design of the exergames were described as good. Especially the simple and clear game design structures were reported to be highly appreciated by patients and to promote good task comprehensibility. Various minor usability issues were reported, including difficulties in the interaction with the exergame training system “Senso” (eg, unintendingly walk off the middle plate without noticing the feedback on the screen), but mainly, usability issues that related to capabilities of older adults with mNCD (eg, limited comprehensibility of the game instructions) were reported. These usability issues need to be considered and addressed when developing a training concept specifically for older adults with mNCD. Nevertheless, it is important to emphasize that these are only minor usability issues, and only minor refinements are required to optimize the exergame experience. This is also illustrated by recent studies showing that exergame-based training programs using the “Senso” are feasible; usable; and widely accepted in different populations including community-dwelling older adults [84], geriatric inpatients [51], and patients with major NCD [52], chronic stroke [85], or multiple sclerosis [86]. Therefore, when designing and developing an exergame-based training concept specifically for older adults with mNCD, these refinements should primarily target the adaptability and individualization of task demands and the optimization of the instruction of the exergames.

Motivators for Training

The motivating factors most frequently described by experts were classified as intrinsic motivators. These were described as being maintained by the captivating character of exergames and promoted by specific game components such as game tasks or designs close to everyday life or with a personal relation or memory, including music or sound effects, animals or plants, landscapes, or colors. In addition, patients were described to be intrinsically motivated by gamification and the feeling of being optimally challenged. From a patient’s perspective, the effectiveness of the training, which helped them achieve their individual success, was clearly the most prominent motivator.

This is consistent with reports in the literature. More autonomous forms of motivation can be promoted by various factors, although these factors may vary depending on the population. For example, a small case-control study compared the motivational factors for using a balance exergame platform between healthy younger and older adults. It was shown that “older adults were more intrinsically motivated by the joy of playing and extrinsically motivated by the perceived health effects (physical and cognitive), with less regard for the in-game rewards” [87]. To provide effective interventions to promote physical activity [88] in patient with NCDs, a new theoretical model has recently been introduced. This theoretical model is based on the review of existing theories that explain behavior change in relation to physical activity in HOA, which were then adapted and integrated to a new theoretical model called the “PHYT in dementia” [88]. In this framework, several additional key elements for promoting behavioral changes in physical activity have been proposed. These consist of self-efficacy, including embarrassment (eg, supervision of activity had a negative impact on engagement in the intervention), personal concerns (eg, fear of falling), and routine (eg, flexible integration of physical activity intervention into daily life regarding place and time of performance), as well as appropriate challenges [88]. A detailed awareness of participant motivators is required, especially for the preference that the routine can be performed at home and at different times during the day [88], because self-determined motivation may be a central aspect of adherence to home-based training programs [89].

Training Goals and Outcomes

The interviewed experts recommended to mainly target cognitive functioning when developing a training concept for older adults with mNCD, while ADLs and mobility, physical capabilities, and psychosocial factors should also be accounted for. This is consistent with the patients’ viewpoint that most frequently reported improving gait and memory as their primary training goals to increase their quality of life.

Similar results have been documented in the literature. According to a survey of patients who completed a multicomponent behavioral intervention for patients with MCI and their caregivers, quality of life was the most important outcome priority for patients with MCI, followed by self-efficacy, depression, basic ADLs, memory-based ADL, anxiety, and memory performance [90].

Exergame and Training Components

The use of exergames as a form of simultaneous-incorporated motor-cognitive training is recommended, which should be prescribed domain-specifically, depending on a patient’s cognitive abilities. Previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have used commercially available exergame systems [76,91-96]
or exergames that were specifically developed for patients with mNCD or MCI [97-100], which comprised sensor-based stepping platforms [94], video camera–based or wireless remote device systems [76,91,95,97,99], or exergames that were controlled using a cycle ergometer or similar [92,93,96,98,100]. The training programs can be classified as simultaneous-additional [92,93,96,98] or simultaneous-incorporated [76,91,94,95,97,99,100] motor-cognitive training that was applied targeting 1 [93,100] or multiple [76,91,92,94-99] neurocognitive domains, including complex attention [76,91,92,94-99], executive functions [76,91,92,94-100], learning and memory [91,93-95,97-99], or visuospatial skills [76,97,99]. Only one of these studies applied training that individually prescribed content on the basis of a patient’s cognitive abilities [96]. However, it has not been performed or reported in a reproducible manner.

Therefore, so far and to the best of our knowledge, 11 studies have been published that investigated exergame-based motor-cognitive training in older adults with mNCD or MCI. Most of these studies designed or used exergames that could be classified as simultaneous-incorporated motor-cognitive training. Incorporating cognitive tasks into motor tasks may be more beneficial for consolidating neuroplasticity [42], because (1) it leads to greater (motor) cognitive improvements, (2) it is closer to daily life situations, (3) no prioritization effects occur, which can be observed in motor-cognitive training with additional cognitive tasks, and (4) multiple sensory systems are stimulated at the same time, which may provide an optimal basis for cognitive processes such as learning [42]. Meta-analytic evidence suggests that simultaneous motor-cognitive training is the most effective type of training for improving cognition in HOA [43,44] and in older adults with mNCD [44-46]. Nevertheless, it remains to be evaluated whether the incorporation of cognitive tasks into exercise or training interventions indeed results in more distinct effects on cognitive performance compared with simultaneous motor-cognitive training with a non–task-relevant secondary cognitive task [42]. Finally, there seems to be room for improvement regarding the domain-specific prescription of the training content, considering a patient’s cognitive abilities and the adaptation and development of exergames specifically for patients with mNCD. This may be especially relevant when considering the large heterogeneity in the clinical symptoms of older adults with mNCD. Remarkably, most previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have used commercially available exergame systems [76,91-96], in which the training content does not specifically target patients with mNCD. This is consistent with the findings of HOA. In a systematic review, Valenzuela et al [80] emphasized that in HOA, most studies used commercially available exergame systems. It was argued that these systems may be difficult to use for those with little or no experience with technology, because these systems often lack clear instructions, present too much graphical information, and have not been designed and developed to provide optimal training components for the target population and aims of the studies in which they were used [80]. In fact, all previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have used exergames with complex 2D or 3D virtual environments [76,91-100]. This may not be optimal because the limitation that such systems may be difficult to use for those with little or no experience with technology could be even more pronounced in patients with mNCD, as these patients are easily distracted and quickly overwhelmed by the task demands. Indeed, according to the recommendations of the interviewed experts, it is beneficial to focus on the aspects that need to be worked on by implementing easily comprehensible and clearly designed exergame tasks and to only present elements that are directly related to the game tasks while avoiding unnecessary graphical information or distractors.

According to the recommendations of the interviewed experts, the training program should be maintained over the long term (preferably ≥12 weeks). A training frequency of 2 to 5 or more training sessions per week was recommended, largely depending on the training location and motivation. In addition, it is recommended to reach a moderate training volume of approximately 150 minutes per week. To reach this training volume, shorter training sessions and a higher training frequency should be applied, because longer training sessions might lead to attentional exhaustion in this group of patients. Therefore, the experts recommended session durations between 15 and 20 minutes up to a maximum of 30 minutes. Previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have prescribed training programs over durations of 5 weeks [97], 6 weeks [91,94,98,100], 12 weeks [96,99], 3 months [93,95], 24 weeks [76], or 6 months [92]. The prescribed training frequency was 1 time per week [76], 2 times per week [94,96,99], 2 to 3 times per week [95], 3 times per week [97,98], 3 to 5 times per week [92,93,100], or 5 times per week [91] with session durations of 15 minutes [99], 18 to 30 minutes [94], 20 to 80 minutes [97], 25 to 30 minutes [91], 20 to 40 minutes [93], 30 to 45 minutes [100], 40 to 45 minutes [98], 45 minutes [92], 60 minutes [96], 90 minutes [76], or not reported [95], resulting in a weekly training volume of 30 minutes [99], 36 to 60 minutes [94], 60 to 200 minutes [93], 90 minutes [76], 90 to 225 minutes [100], 100 to 145 minutes [97], 120 minutes [96], 120 to 135 minutes [98], 125 to 150 minutes [91], 135 to 225 minutes [92], or not reported [95]. Therefore, most of these studies prescribed a training volume that was in line with the recommendations of the experts in this study. However, the session durations often exceeded the experts’ recommendations, whereas the training frequency was lower than recommended. To avoid attentional exhaustion of the patients during training, future training concepts might consider prescribing shorter session durations while increasing the training frequency to achieve a similar training volume per week. This might actually improve the effectiveness of the intervention because higher training frequencies have already been shown to promote the effectiveness of physical (ie, ≥4 times per week) [101] and cognitive training (ie, ≥3 times per week) [102], while shorter session durations (ie, ≤30 minutes) [101] of physical exercise have been shown to exert more pronounced training effects. These findings might also apply to simultaneous motor-cognitive training. A meta-analysis revealed that training frequency is a significant moderator of the effects of physical and motor-cognitive training interventions on cognitive functioning, favoring higher training frequencies (≥5 times per week) in a mixed population of HOA and patients.
with mNCD [103]. Finally, a high training frequency (approximately 5 times per week) with short session durations (approximately 20 minutes) would also match the preferences of the interviewed patients in this study.

The experts reported that the training should preferably be individually carried out at the patients’ homes, not only because it represents a known environment that makes patients feel more secure and represents a less-confronting environment for them (because they do not have to hide their impairments from others when training alone), but also to allow higher training frequencies. Nonetheless, to ensure that training in patients’ homes is feasible, multiple factors need to be considered. For example, improvements in game instructions are required, a handrail or similar needs to be made available to allow safety support during training, and technical problems must be avoided. In addition, a guided familiarization period and part-time supervision or support from a care professional or partner should be integrated to make the transfer to home-based exergaming easier. Previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have administered individual [91] or group-based [76,96,99] training sessions, and the training setting (ie, individual vs group sessions) has not been clearly reported [92-95,97,98,100].

The training sessions were conducted at the hospital [91], in a nursing home [94], at day-care centers or memory clinics [99], at a centrally located church [76], at patients’ homes [93,100], or the training location was not clearly reported [92,95-98]. The training sessions were supervised by a therapist [91,94,96], or supervision was not reported [76,92,93,95,97-100]. Consistent with summarized previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI, most cognitive training programs to date have also been conducted in group sessions [104]. However, most of our interviewed patients clearly stated that they would prefer to train individually at home or with the support of a care professional or partner. Therefore, it might be worthwhile to put more effort into designing and developing exergames that can be used individually at home. This would possibly also reduce the barriers of patients with mNCD to engage in exergame-based training programs in the long term.

Regarding training demands, the experts recommended focusing on game complexity to ensure a challenging but feasible cognitive demand. Physical exercise intensity should be maintained at a light to moderate level. To allow individualization of the cognitive demand in training, two main aspects should be considered: (1) task type (ie, choice of exergames to individually focus on neurocognitive functioning) and (2) task demands. To allow individualization of task demands, the following factors should be varied based on the experts’ recommendations: (1) stability support (use of handrail with both hands, one hand, or no support), (2) stepping direction, (3) game choice and tasks included, (4) game duration, or (5) game speed. Previous studies applying exergame-based motor-cognitive training in older adults with mNCD or MCI have applied relatively effortful, high cognitive demands [92], low [97] to moderate [96-98] physical exercise intensities, or have not reported the physical [76,91-95,99,100] or cognitive [76,91,93-100] exercise load or training progression in a clearly reproducible way. This exemplifies the fact that the optimal cognitive load for motor-cognitive training remains unknown. To the best of our knowledge, there has only been 1 meta-analysis to date that compared the effects of training interventions on cognitive functioning in relation to different task complexities and found no difference between simple and complex cognitive games [105]. Therefore, further investigations are needed to identify the optimal cognitive training demands and optimize the monitoring and progression of training programs. For physical exercise intensity, the recommendations of the interviewed experts are in line with those of previous studies applying exergame-based motor-cognitive training in older adults with mNCD. This also matches the analysis of the moderating variables of the training parameters that influence the effectiveness of the interventions. Based on meta-analytic results from motor-cognitive training in older adults with mNCD, moderate physical training intensity [45] has been shown to be most effective in improving cognitive function. Finally, moderate physical exercise intensity would also match the preferences of the patients interviewed in this study.

Implications for Research

Our findings serve as a basis for user modeling, determination of therapeutic needs, and definition of a set of requirements for the game design and development of novel exergame-based training concepts. To increase the probability that the resulting training will be deemed feasible in future clinical practice, these considerations should be integrated to guide the decision process for the most suitable exergame design and intervention components when developing novel exergames and exergame-based training concepts.

Limitations

The outcomes of this qualitative study must be interpreted with some caution, considering the following limitations. First, none of the interviewed patients with mNCD belonged to the clinical subtypes of mild frontotemporal NCD or mNCD with Lewy bodies. Depending on the clinical subtypes and the associated clinical pictures of the patients, different findings may have emerged from patient interviews. However, a substantial fraction (ie, ≥60%) of mild or major NCD is attributable to Alzheimer disease, whereas mild vascular NCD is the second most common cause of NCD after Alzheimer disease; frontotemporal NCD only accounts for approximately 5% of cases [9]. Therefore, the included study population appeared to be representative of these clinical subtypes. Second, owing to difficulties in recruiting patients, those screened for MCI according to predefined criteria were recruited in addition to patients with a clinical diagnosis of mNCD, which increased the heterogeneity of the study population. By contrast, in our project, we aimed to develop an individualized exergame-based training concept not only to treat clinically diagnosed patients with mNCD but also to prevent progression to dementia in individuals at risk who might not have been diagnosed (yet). Third, owing to the COVID-19 pandemic, all focus group sessions were held as web-based meetings. Face-to-face focus group sessions might have promoted livelier exchanges and may have led to additional insights.
Conclusions
The psychosocial consequences of patients’ self-perceived cognitive deterioration may be more burdensome than the cognitive changes themselves. Older adults with mNCD prefer integrative forms of training (such as exergaming) and are primarily motivated by enjoyment or fun in exercising and the effectiveness of the training. Putting the synthesized perspectives of training goals, settings, and requirements for exergames and training components into context, our considerations point to opportunities for improvement in research and rehabilitation, either by adapting existing exergames to patients with mNCD or by developing novel exergames and exergame-based training concepts specifically tailored to meet patient requirements and needs.

Acknowledgments
The authors would like to thank the Synapsis Foundation—Dementia Research Switzerland (grant 2019-PI06) for funding this project. The authors would like to thank all the participants in this study for their participation and valuable contributions to this project.

Data Availability
The data supporting the findings of this study are available from the corresponding author (PM) on reasonable request.

Authors’ Contributions
PM was responsible for the conception and protocol development of this study under the supervision of EDdB. MAO contributed to the study conception. PM was responsible for the recruitment of participants, data collection, data analysis, and writing of the manuscript. Data coding and analysis was cross-checked by MAO. All authors have contributed to the revision of the manuscript. All authors have read and approved the submitted version.

Conflicts of Interest
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest. Dividat AG was asked to suggest suitable participants for the expert focus group discussions by providing a contact list of experts and health care professionals, as the company is well connected with institutions for geriatric populations, physiotherapies, and rehabilitation clinics in Switzerland. Therefore, we were able to identify experts and health care professionals who experienced exergame training with older adults with mild neurocognitive disorder. Experience was preferred with the exergame training system “Senso (Flex)” or similar. In addition, the founder of Dividat AG was partaking in one of the focus group discussions as an industry representative in line with the Multidisciplinary Iterative Design of Exergames framework. Dividat AG had no other role in the study and did not play any role in the design and conduct of the study; they also did not play any role in the data analyses, interpretation, or decision to submit results.

Multimedia Appendix 1
Interview guides.

References


Abbreviations
- ADL: activities of daily living
- CCT: computerized cognitive training
- DSM-5: Diagnostic and Statistical Manual of Mental Disorders 5th Edition
- HOA: healthy older adults
- MCI: mild cognitive impairment
- MIDE: Multidisciplinary Iterative Design of Exergames
- mNCD: mild neurocognitive disorder
- MTT: medical training therapy
- NCD: neurocognitive disorders
- PT: physical therapy

©Patrick Manser, Manuela Adcock-Omlin, Eling D de Bruin. Originally published in JMIR Serious Games (https://games.jmir.org), 05.01.2023. 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 work on https://games.jmir.org, as well as this copyright and license information must be included.
Differences in Brain Activity and Body Movements Between Virtual Reality and Offline Exercise: Randomized Crossover Trial

Hee Jin Kim1*, MD, PhD; Jea Woog Lee2*, PhD; Gangta Choi1, MSc; Junghoon Huh3, PhD; Doug Hyun Han1, MD, PhD

1Department of Psychiatry, College of Medicine, Chung-Ang University, Seoul, Republic of Korea
2Department of Information & Technology in Sport, College of Sports Science, Chung-Ang University, Anseong-si, Gyeonggi-do, Republic of Korea
3Department of Human Motor Behavior, College of Sports Science, Chung-Ang University, Anseong-si, Gyeonggi-do, Republic of Korea
*these authors contributed equally

Corresponding Author:
Doug Hyun Han, MD, PhD
Department of Psychiatry
College of Medicine
Chung-Ang University
Heuk Seok Ro 84
Seoul, 06974
Republic of Korea
Phone: 82 2 6299 3132
Fax: 82 2 6298 1508
Email: hduk70@gmail.com

Abstract

Background: Virtual reality (VR) has been suggested to be effective at enhancing physical exercises because of its immersive characteristics. However, few studies have quantitatively assessed the range of motion and brain activity during VR exercises.

Objective: We hypothesized that 3D immersive VR could stimulate body movement and brain activity more effectively than standard exercises and that the increased range of motions during 3D immersive VR exercises would be associated with orbitofrontal activation.

Methods: A randomized crossover trial was conducted to compare exercises with and without VR. A total of 24 healthy males performed the same motions when exercising with and without 3D immersive VR, and the recorded videos were used for motion analysis. Hemodynamic changes in the prefrontal cortex were assessed using functional near-infrared spectroscopy.

Results: There were significant differences in the total angle ($z=-2.31; P=0.02$), length ($z=-2.78; P=0.005$), calorie consumption ($z=-3.04; P=0.002$), and change in accumulated oxygenated hemoglobin within the right orbitofrontal cortex ($F^1.94=9.36; P=0.003$) between the VR and offline trials. Hemodynamic changes in the right orbitofrontal cortex were positively correlated with the total angle ($r=0.45; P=0.001$) and length ($r=0.38; P=0.007$) in the VR exercise; however, there was no significant correlation in the offline trial.

Conclusions: The results of this study suggest that 3D immersive VR exercise effectively increases the range of motion in healthy individuals in relation to orbitofrontal activation.

Trial Registration: Clinical Research Information Service KCT0008021; https://cris.nih.go.kr/cris/search/detailSearch.do/23671

(JMIR Serious Games 2023;11:e40421) doi:10.2196/40421

KEYWORDS
virtual reality exercise; near-infrared spectroscopy; spectroscopy; hemodynamic; blood flow; hemoglobin; brain; prefrontal cortex; orbitofrontal cortex; immersion; virtual reality; VR; exercise; range of motion; physical activity; fitness; motion; movement; randomized; calorie
Introduction

Virtual reality (VR) is an artificially created sensory experience using digital technology, which provides users with a subjective sense of "being there" [1]. The format of VR, referring to the structure of the information displayed, can be either 2D or 3D multimedia. The main difference between the 2 formats is the level of interactivity, with 3D environments allowing changes in the user’s point of view, navigation, and interaction with objects and people. The display devices, which are the technological equipment used to visualize the formats, are classified based on the level of immersion they provide: nonimmersive, semi-immersive, and immersive. Nonimmersive systems are simple devices that use a single screen, such as a desktop PC. Semi-immersive systems, such as the Cave automatic virtual environment and the Powerwall display, provide a stereo image of an environment using a projection linked to the observer’s position. Immersive systems, such as a head-mounted display (HMD), isolate the user from external world stimuli [2]. An HMD gradually follows the users' movements, quantifies the position and direction of articles in reality, and creates a corresponding stereoscopic view. Users perceive themselves to be situated in a virtual environment and are temporarily unable to recognize behaviors and objects in the real environment [1].

VR is expected to have enormous potential in fields such as education, training, and entertainment [3]. In particular, the application of VR technology is promising in the field of sports, as it can overcome human limitations due to technical innovations, thus allowing for the development of effective training methods [4].

VR application in sports can increase motivation, enjoyment, and physical performance [5]. This may be due to the unique characteristics of VR, such as immersion, interaction, and infinite degrees of freedom in spatial and temporal situations [6,7]. Fox and Bailenson [8] showed that physical activity could be increased through a reinforcement mechanism using immersive VR technology. Jones et al [9] demonstrated that immersion in VR exercises was positively correlated with motivation, pleasure, and enjoyment after exercises. In addition, participants can interact with objects in a stereoscopic visual space without spatial limitations, which can be manipulated in various ways, depending on individual needs [10]. Athletes can use VR exercises to learn new sports strategies and compete with their virtual partners [10]. By using the unlimited spatial and temporal boundaries of VR, injury-prone skills can be safely trained [11]. For example, VR allows ski jumps and downhill skiing in a safe environment [11].

The shortcomings of physical exercise using an immersive VR have been attributed to the use of HMDs. Exercising while wearing an HMD can be uncomfortable due to sweating and head bumps [12]. In addition, HMDs can be impractical for movement accuracy and speed. They can also be potentially dangerous (eg, when the device blocks the vision of the moving treadmill while running [12]). Furthermore, interactive VR display systems linked to various eye-movement systems can cause motion sickness [13].

VR exercises are effectively used in the rehabilitation of patients who have lost motor skills [5,14]. They are also used to increase the physical fitness of patients with chronic medical conditions, such as hemodialysis or dementia, and developmental disorders, such as intellectual disability or autism [15,16]. Most previous studies on VR exercises have been conducted on patient groups [5,14-16]. However, VR exercises are increasingly used by athletes and other sports-related professionals, as they enable individualized training with real-time feedback without coaches [11,12]. Moreover, VR exercises have recently been widely used in the public sphere to promote physical fitness [5]. When applied to healthy individuals, VR exercise can increase exercise speed, heart rate, balance, motivation, enjoyment, adherence, executive function, and neuroplasticity [17-20]. Although many studies have suggested the qualitative effectiveness of VR exercise, few have quantitatively assessed its effectiveness [12].

Previous studies have suggested that the brain regions of interest in response to VR and physical exercises are the dorsolateral prefrontal cortex (DLPFC), ventrolateral prefrontal cortex (VLPFC), and orbitofrontal cortex (OFC) [9,21,22]. Of these 3 major prefrontal cortical regions, a functional near-infrared spectroscopy (fNIRS) study showed that VR stimuli activated the OFC [21]. The OFC has been associated with emotional and behavioral regulation, decision-making, maintenance of behavioral flexibility, and processing of anticipated rewards and punishments [23]. Inappropriate functioning of the OFC can lead to disinhibition, preservation, and impulse control problems [24]. Owing to the functions of rewards and decisions, the OFC can be associated with VR-induced immersion and flow [8,25].

This study aimed to investigate whether immersive VR technology can enhance physical exercise, in terms of the range of motion and brain activity, in healthy individuals. Kinematics and resting-state fNIRS methodologies were adopted to demonstrate the quantitative effectiveness of 3D immersive VR exercise. Kinematics is useful for objectifying human movements and comparing different subjects or situations, such as before and after treatment or training in various fields (eg, sports science) [26]. For example, knee and hip angles at foot strikes during running have been quantified using 2D video recording and motion analysis software [27]. Vertical jump height, frequently used as an indirect measure of muscle power in the lower extremities, was also estimated using the video method [28]. In VR rehabilitation studies, the difference in ranges of motion before and after treatment was measured to assess improvement [29]. fNIRS is a noninvasive functional brain imaging tool that measures cortical hemodynamic activity [30]. Near-infrared light penetrates the brain, and oxygenation changes in regional cerebral blood flow are recorded [31]. fNIRS has the advantages of safety, low cost, portability, tolerance to motion artifacts, and good temporal and spatial resolution, overcoming the shortcomings of conventional brain imaging modalities, such as electroencephalography and functional magnetic resonance imaging (fMRI) [32,33]. fNIRS has been used in various fields, including sports medicine, neuroscience, behavioral science, clinical studies, and brain-computer interfaces [30].
We hypothesized that 3D immersive VR exercises would be more effective than offline physical exercises in increasing the range of motion and calorie consumption during exercise. In addition, we hypothesized that the effectiveness of 3D immersive VR exercises is associated with brain activation within the OFC before and after exercise.

Methods

Sample Size
The sample size was calculated using G*power software (Heinrich-Heine-Universität Düsseldorf) [34] based on a statistical test using repeated measures ANOVA. A type 1 error of .05 and a statistical power of 0.9 were used. The correlation among repeated measures was set conservatively at 0. Based on previous similar studies, an effect size of 0.25 was estimated. Based on these calculations, 22 participants were required. After estimating a 10% dropout rate, the required total sample size was set at 24 participants.

Participants
Through flyer advertisements, 24 healthy males were recruited for this study from Chung-Ang University. The inclusion criteria were as follows: (1) male sex, (2) age 20 to 29 years, and (3) no psychiatric or medical illness. The exclusion criteria were as follows: (1) history of head trauma; (2) history of substance abuse, including alcohol, tobacco, and drugs; and (3) IQ<80. The participants were instructed not to consume any food, caffeine, or alcohol 3 hours prior to participation in the study.

Ethics Approval
The institutional review board of Chung-Ang University approved this research protocol (1041078-201908-HRSB-231-01). Written informed consent was obtained from all participants.

Study Procedure
Overview
This study was a randomized crossover trial that compared responses to 3D immersive VR and offline exercises. The participants were blinded to the study hypotheses. Each participant underwent 1 3D immersive VR trial and 1 offline trial in a randomly assigned order at 10-minute intervals. Each session consisted of (1) a 2-minute practice to become familiarized with each exercise; (2) a 2-minute pre-exercise resting-state fNIRS measurement; (3) a 5-minute exercise, either 3D immersive VR or offline; and (4) a 3-minute postexercise resting-state fNIRS measurement (Figure 1). This trial was conducted in a curtained, dimly lit gym at a temperature of 23 °C with minimal ambient noise to control external noises as much as possible. The participants were instructed not to speak or move their heads during the fNIRS measurement to maintain good tissue-electrode contact. A 10-minute interval was given between sessions to return noncortical hemodynamic variables, such as skin blood flow and middle cerebral artery blood flow, to baseline levels.

Both VR and offline exercises consisted of 40 identical movements derived from Pilates. Pilates was chosen because it is a beneficial exercise method for healthy adults that enhances flexibility, muscular activity, and coordination [35]. In both exercises, the sequence was in the order of 8 types of simple behaviors involving just the body, arms, or legs; 16 types of 2 complex behaviors involving the body plus arms or legs; and 16 types of 3 complex behaviors involving 3 parts of the body, arms, and legs. The reason for exercising in this order was to gradually increase the number of muscle groups involved, starting from warming up. A resistance band (TheraBand) was applied to each participant in both sessions to maintain adequate exercise intensity and prevent less controlled behaviors. Pilates exercise with a resistance band was reported to increase muscle activity by more than 50%, whereas Pilates exercise without a resistance band increased muscle activity by only 20% [36]. Before the VR exercise, the research staff provided brief instructions on the 3D immersive VR system and environment and asked the participants to report whether they experienced any discomfort while wearing the HMD, such as head bumps, sweating, or motion sickness. The research staff checked for any obstacles the participants might encounter during the VR exercise and monitored and provided immediate assistance to the participants during the exercise in case of danger. With the help of the research staff, the participants were positioned 2 m horizontally away from the VR device, wearing the HMD and holding resistance bands in both hands at shoulder level. In this study, the Vive Cosmos Elite (High Tech Computer Corp) and Kinect V1 (Microsoft) were used as the HMD and motion sensors, respectively. A 3D map–based immersive VR software system that allows participants to interact with the VR...
environment was developed using Unity3D Engine (Unity Technologies). The Grand Canyon landscape was designed to enable participants to fly between canyons without crashing by performing the aforementioned movements in sequence. During the VR exercise, participants were asked to fly over the canyon without bumping into it, following visual and audio guidance. The direction in which to move, whether to accelerate or decelerate, and flight speed were displayed on the VR screen in real time, serving as visual instructions (Figure 2). When the participants tilted their upper body right, left, forward, or backward in the virtual space, they rotated right, left, upward, or downward, respectively. They could accelerate or decelerate by lifting or lowering their right arm.

In contrast, during the offline exercise, the participants performed movements in a standard manner following a demonstration by a professional instructor. In both sessions, audio guidance, such as “lean right and raise your left arm” were provided.

The research staff helped the participants accurately perform each movement in both sessions. They provided immediate feedback for incorrect movements and high-risk situations. The movements performed by the participants in the VR and offline conditions were considered identical since the audio guidance, use of the resistance band, and real-time feedback on behavior by the research staff were the same for all participants.

**Figure 2.** A representation of the virtual environment.

---

**Motion Analysis to Measure Motion Angle, Length, and Retention Time**

Participants’ movements during the VR and offline exercises were recorded using an iPhone 11 camera (Apple). The participants were asked to stand at the designated site 1.5 m away from the camera to maintain a constant camera angle. The recorded video was analyzed using the open-source video analysis software Kinovea (version 0.8.15; Kinovea Open-Source Project) [21]. Kinovea determines the distance between the camera and the recorded object by measuring the subject passing in front of the camera [21]. It has various analysis and measurement tools for adding annotations, drawing lines, and calculating distances and angles [37] and has been used in numerous sports and clinical studies [22-24]. Previous studies have shown the reliability and validity of the software at distances of up to 5 m with an angle of 90° to 45° [21].

In the motion analysis, eight types of motion angles were measured: (1) right waist angle, (2) left waist angle, (3) right body angle, (4) left body angle, (5) right leg angle, (6) left leg angle, (7) forward body angle, and (8) backward body angle. Additionally, four types of movement lengths were measured: (1) right arm movement length, (2) left arm movement length, (3) right leg movement length, and (4) left leg movement length (Figure 3). The total angle and movement length for simple behaviors, 2 complex behaviors, and 3 complex behaviors were calculated using each angle and movement length involved in each behavior. For example, for simple behaviors, 8 angles and 4 lengths were used to calculate the total angle and movement length. Two complex behaviors consisting of 16 movements, 24 angles, and 16 lengths were measured to calculate the total angle and movement length. Finally, 3 complex behaviors consisting of 16 movements, 24 angles, and 32 lengths were measured to calculate the total angle and movement length (Multimedia Appendix 1).
Figure 3. Movement angle and length. From top: Rt (Lt)-WA: right (left) waist angle; Rt (Lt)-BA: right (left) body angle; Rt (Lt)-LA: right (left) leg angle; FBA (BBA): forward (backward) body angle; Rt ArML: right arm movement length; Lt ArML: left arm movement length; Rt lgML: right leg movement length; Lt lgML: left leg movement length.

Comparison of Calories Using the Metabolic Equivalent of Task Formula and Motion Retention Time
Calorie consumption during each session was calculated using each participant’s weights, motion retention time (from the start to the end of the movement), and the metabolic equivalent of task (MET) of the Pilates motion, according to the MET formula [38]. The MET formula has been adopted in various clinical studies, and its reliability has been demonstrated through meta-analyses [39,40]. MET is defined as the level of intensity of various physical activities, including daily activities and low-intensity to high-intensity exercises. The Pilates movements performed by the participants in this study were rated as 3.5 MET.

Hemodynamic Changes in the Frontal Cortex
Hemodynamic changes within the prefrontal cortex were assessed using a high-density fNIRS device (NIRSIT, OBELAB Inc). The curved panel of the NIRSIT has 24 laser diodes (sources) emitting 2 wavelengths (780 nm and 850 nm) of light and 32 photodetectors with a sampling rate of 8.138 Hz. The unit distance between the source and photodetectors was 15 mm. In this study, only the 30-mm channels were analyzed, since 30 mm is the most suitable sensor-detector separation distance for measuring cortical hemodynamic changes.

The fNIRS data were processed using the analysis toolbox provided by the fNIRS device manufacturer. The light signals in each wavelength within the 48 channels were filtered using a band-pass filter (between 0.0 Hz and 0.1 Hz) to reduce noise due to external light and body movements. Data derived from channels representing low-quality information (signal-to-noise ratio<30 dB) were not included in the hemodynamic analysis set to prevent misinterpretation. Relative hemodynamic changes during exercise were calculated using the modified Beer-Lambert law [41]. The accumulated oxygenated hemoglobin (accHbO2) values represent the activation of the prefrontal cortex during a resting state. Although both oxygenated and deoxygenated hemoglobin data could be obtained, previous studies have revealed that oxygenated hemoglobin has superior sensitivity and signal-to-noise ratio; therefore, only oxygenated hemoglobin data were used for this analysis [26,28].
Figure 4. The channel configuration.

Statistical Analyses
For total behaviors, simple behavior, 2 complex behaviors, and 3 complex behaviors, the differences in total angles, total movement length, and consumed calories were compared using the Mann-Whitney \( U \) test. The statistical significance of the total angle and total movement length in the total behaviors was set at 0.02 (0.05/3, angle, length, and calories). In the post hoc test, the statistical significance of the total angle and total movement length in the simple behaviors, 2 complex behaviors, and 3 complex behaviors was set at \( P=.02 \) (0.05/3, angle, length, and calories). Between the VR and offline situations, the difference in the change in accumulated oxygenated hemoglobin (\( \Delta \text{accHbO}_2 \)) from baseline and exercise was compared using repeated measures ANOVA. The statistical significance was set at \( P=.006 \) (0.05/8, 8 brain regions). Correlations between the differences in accHbO2 within the right OFC, total angles, and total movement lengths were assessed using Pearson correlation coefficient. Statistical significance was set at \( P=.03 \) (0.05/2, angle and length). Statistical analyses were performed using SPSS software (version 24.0; IBM Corp).

Results

Differences in Movement Including Angle and Length Between Offline and VR Trials During Exercise
During exercise, there were significant differences in the total angle (\( z=-2.31; P=.02 \)) and length between the VR and offline trials (\( z=-2.78; P=.005 \)). The VR group showed a 13.3% wider total angle range and an increase of 14.1% in the total movement length compared with the offline group. In the post hoc analysis, the VR group moved significantly more in simple and complex behaviors, except for the total angle in 2 complex behaviors, compared to the offline group (Table 1).
Table 1. Differences in angle and distance between the VR\(^a\) and offline trials\(^b\).

<table>
<thead>
<tr>
<th>Types of behavior and parameters</th>
<th>VR trial, mean (SD)</th>
<th>Offline trial, mean (SD)</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total behavior</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle total (°)</td>
<td>1359.39 (125.55)</td>
<td>1200.26 (195.68)</td>
<td>(z=-2.31; P=.02)</td>
</tr>
<tr>
<td>Length total (cm)</td>
<td>2568.20 (342.38)</td>
<td>2250.82 (382.20)</td>
<td>(z=-2.78; P=.005)</td>
</tr>
<tr>
<td>Calorie total (kcal)</td>
<td>9.36 (2.60)</td>
<td>5.79 (3.01)</td>
<td>(z=-3.04; P=.002)</td>
</tr>
<tr>
<td><strong>Simple behavior (trunk leaning over)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle total (°)</td>
<td>172.30 (19.2)</td>
<td>147.8 (27.64)</td>
<td>(z=-2.74; P=.005)</td>
</tr>
<tr>
<td>Length total (cm)</td>
<td>224.93 (34.3)</td>
<td>197.70 (31.9)</td>
<td>(z=-2.53; P=.01)</td>
</tr>
<tr>
<td>Calorie total (kcal)</td>
<td>2.46 (0.55)</td>
<td>1.18 (0.69)</td>
<td>(z=-4.30; P&lt;.001)</td>
</tr>
<tr>
<td><strong>2 complex behaviors (trunk leaning over + lifting arm or leg)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle total (°)</td>
<td>449.26 (118.57)</td>
<td>415.87 (69.17)</td>
<td>(z=-1.14; P=.27)</td>
</tr>
<tr>
<td>Length total (cm)</td>
<td>966.09 (118.57)</td>
<td>860.06 (107.01)</td>
<td>(z=-3.10; P=.001)</td>
</tr>
<tr>
<td>Calorie total (kcal)</td>
<td>4.42 (1.35)</td>
<td>2.95 (1.58)</td>
<td>(z=-2.55; P=.01)</td>
</tr>
<tr>
<td><strong>3 complex behaviors (trunk leaning over + lifting arm + lifting leg)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angle total (°)</td>
<td>737.83 (63.61)</td>
<td>636.55 (112.77)</td>
<td>(z=-2.72; P=.006)</td>
</tr>
<tr>
<td>Length total (cm)</td>
<td>1377.18 (202.88)</td>
<td>1193.05 (125.55)</td>
<td>(z=-2.69; P=.006)</td>
</tr>
<tr>
<td>Calorie total (kcal)</td>
<td>2.47 (0.86)</td>
<td>1.65 (0.97)</td>
<td>(z=-2.63; P=.009)</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.
\(^b\)\(P<.02\) was considered to be significant.

**Differences in Consumption Between Offline and VR Trials During Exercise**

During exercise, there were significant differences in the total calories between the VR and offline trials (\(z=-3.04; P=.002\)). During exercise, the VR group consumed 61.7% more calories than the offline group. In the post hoc analysis, the VR group showed significant calorie consumption with 2 complex and 3 complex behaviors compared to the offline group (Table 1).

**Differences in \(\Delta \text{accHbO2} \) Between Offline and VR Trials During Exercise**

During exercise, significant differences were observed in \(\Delta \text{accHbO2} \) between the offline and VR trials within the right OFC (\(F^{1,94}=9.36; P=.003\)) but not within the left VLPFC (\(F^{1,94}=5.69; P=.02\)), left DLPFC (\(F^{1,94}=0.06; P=.81\)), right DLPFC (\(F^{1,94}=0.02; P=.89\)), right VLPFC (\(F^{1,94}=0.06; P=.81\)), left OFC (\(F^{1,94}=0.01; P=.91\)), left frontopolar prefrontal cortex (\(F^{1,94}=0.05; P=.89\)), or right frontopolar prefrontal cortex (\(F^{1,94}=0.21; P=.65\)) (Figure 5).
Correlation of ΔaccHbO2 With Total Angle and Movement Length

In all participants, the ΔaccHbO2 within the right OFC was positively correlated with the total angle ($r=0.36; P=.005$). However, it was not significantly associated with the total movement length ($r=0.28; P=.40$). In the VR group, the ΔaccHbO2 within the right OFC was positively correlated with total angle ($r=0.45; P=.001$) and total movement length ($r=0.38; P=.007$). In the offline group, the ΔaccHbO2 was not significantly correlated with total angle ($r=0.28; P=.40$) or total movement length ($r=0.29; P=.37$) (Figure 6).
Figure 6. Correlation between accumulated oxygenated hemoglobin in the orbitofrontal cortex (OFC) and movement change. $P<.025$ was considered to be significant. Oxy-Hb: oxygenated hemoglobin.

User Experience During VR Exercise
Of the 24 participants, 3 (12.5%) reported motion sickness, and 8 (33.3%) reported sweating after wearing the HMD device; however, none reported HMD slipping or severe discomfort.

Discussion

Principal Results
In this study, VR exercise increased the range of motion, calorie consumption, and brain activity within the OFC compared with offline exercise. In addition, increased brain activity within the right OFC was correlated with an increased range of motion.

Improvement of Movement Range and Calorie Consumption in Response to Immersive VR Stimuli
Our findings are consistent with previous studies on various patient groups. However, no studies have reported changes in the range of motion due to VR exercise in healthy individuals, as in this study [5]. VR exercises have been suggested to improve the range of motion in patients undergoing rehabilitation, both in the short term [42,43] and in the long term [14,44]. Immersive VR physical therapy significantly reduces pain and enhances the range of motion in patients with burn injuries (immediate effect) or chronic frozen shoulders (chronic effect) compared with standard therapies [14,42]. In addition, a 4-week VR exercise program in patients who underwent a stroke was more effective than a conventional program in improving the active range of motion and other scores, such as the Fugl-Meyer assessment score, Wolf motor function test score, and modified Barthel Index [44]. Children with cerebral palsy also reported a greater range of motion, better motion control, and greater interest as an acute effect after performing VR exercises than after conventional exercises [43]. In addition, the results of this study are in line with those of previous studies showing that participants who engaged in VR exercises had higher heart rates and burned more calories than those who performed standard exercises [45]. Therefore, VR has the potential to enhance physical exercises, and it may be a viable complement or alternative to traditional weight loss programs [46].

The larger range of motion and caloric consumption in the VR exercise may be associated with immersion. Immersion can be divided into 3 forms: sensory, imaginative, and challenge-based [47]. Sensory immersion is associated with audiovisual stimulation, and imaginative immersion is related to absorption in the narrative or identification with a character (eg, feelings of empathy and atmosphere). Challenge-based immersion is related to the balance between the activity’s demands and an individual’s motor and mental skills [48]. Challenge-based immersion is very close to the “flow” experience described by Csikszentmihalyi [49]. Immersion and flow share common features such as concentration, distortion of time perception, loss of self-awareness, and intrinsic motivation toward an activity [49]. The flow of physical exercise is related to peak performance [50]. The flow experienced during exercise is an autotelic experience with total concentration, merging of action and awareness, and the paradox of control [51]. In this respect,
we cautiously suggest that VR exercise could be more immersive and helpful in achieving peak performance than exercise alone since sensory and imaginative components are added to the challenge-based immersion of physical exercise.

Increased Brain Activity Within OFC in Response to VR Stimuli

Another important finding of our study was that there were significant differences in brain activation within the right OFC between offline and exercise groups. To date, only a few studies have specifically investigated orbitofrontal activity in response to VR exercises. Most studies have focused on prefrontal activation caused by VR stimuli. Mao et al [52] suggested that VR training activates the prefrontal cortex and improves spatial orientation and motor function. However, recent studies have suggested that the OFC is one of the candidate brain regions that respond to VR [29]. Landowska et al [21] reported hemodynamic changes within the DLPFC and OFC in response to VR therapy for acrophobia. Dong et al [53] suggested that the activation of the OFC in response to VR tasks was greater than that observed in response to slide-based (control) tasks. Moro et al [54] reported that oxygenation in the OFC was increased when an incremental swing balance task in an immersive VR environment was performed. One possible explanation for this is that VR-induced immersion stimulates reward circuits involving the OFC. The OFC is responsible for sensory integration, regulation of visceral responses, learning, prediction, and decision-making for reward and affective values [55,56]. In addition, visual stimulation and immersion in VR are robustly involved in reward circuits [57].

The results of the correlation study between ΔHbO2 within the right OFC and the range of motion during VR and offline exercise showed that an increase in the range of motion and right OFC activation was correlated only in the VR group but not in the offline group. Thus, we suggest that immersion and visual stimulation of 3D immersive VR distracts the user from unpleasant interoceptive sensations caused by physical exercise. It is suggested that distraction caused by VR can reduce the perception of physical discomfort through an intercortical modulation of the anterior cingulate and OFC [9,58,59].

Additionally, the functional hemispheric differences shown in the results of this study are in line with a previous study that suggested that the more pronounced the shift in brain activity to the right hemisphere, the more flow experienced by elite tennis players [60]. Less synchronized brain activity in the left hemisphere and more coherent brain activity in the right hemisphere have been suggested to reflect less interference from irrelevant verbal-analytical processes with motor control mechanisms [60].

Strengths

To the best of our knowledge, this is one of the earliest studies to quantitatively evaluate the effects of 3D immersive VR exercise on brain hemodynamics in healthy individuals. Each participant’s movement was measured using professional video analysis software, unlike other studies based on subjective performance improvement or satisfaction. Another strength of this study is that we directly compared performance in 3D immersive VR and real-world environments while performing the same activities. This comparison is important to reveal how 3D immersive VR influences exercise performance and brain activity. Finally, this study presents the potential for developing smart headsets by integrating wearable devices and connected technologies, such as VR, brain activity measurement, and motion tracking. Although these were performed with different devices in our study, using an integrated smart headset could enable effective exercise and immediate checking of the exercise amount and the consequent changes in brain activity. These wearable technologies and connected solutions with VR capabilities are already widely adopted in the workplace to promote occupational safety, productivity, and worker health. Examples include the Xsens MVN Animate and Xsens MTw Awinda for body movement tracking and the Muse 2 Melon’s headband and IMEC’s wireless electroencephalogram headset for brain wave sensing [61].

Limitations

This study had some limitations. First, due to the small number and gender disparity of the participants, the results cannot be generalized. In a follow-up study on VR exercise, scaling up the sample size by including virtual participants could be considered. However, this was not possible in this study due to the crossover design comparing VR and offline exercise. Second, there were limitations in the adequate pixel rate, capture rate, maximum aperture, depth of field capture, and frame capture rate of the iPhone 11, which was used as a video recording device. Third, since the fNIRS device used in this study can only measure brain activity within the prefrontal lobe, activity in other cortical regions and deep brain areas was not measured. Further studies using other functional imaging devices, such as fMRI, are required to determine the activity of the entire brain. Fourth, we did not employ additional physiological measures. The use of systemic physiologically augmented fNIRS in future studies may yield more meaningful results. Finally, a standardized questionnaire to check the inclusion and exclusion criteria and to assess motion sickness after VR exercise was not used in this study.

Conclusions

This study confirmed that 3D immersive VR exercise is more effective at increasing the range of motion and calorie consumption during exercise in healthy individuals via activation within the OFC.
Conflicts of Interest
None declared.

Editorial Notice
This randomized study was only retrospectively registered. The authors regret that the trial could not be registered prospectively due to an oversight in the process. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials. 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
Performances for simple and complex behaviors.
[DOCX File, 17 KB - games_v11i1e40421_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 97 KB - games_v11i1e40421_app2.pdf]

References


Abbreviations

- accHbO2: accumulated oxygenated hemoglobin
- DLPFC: dorsolateral prefrontal cortex
- fMRI: functional magnetic resonance imaging
**fNIRS**: functional near-infrared spectroscopy

**FPC**: frontopolar cortex

**HMD**: head-mounted display

**MET**: metabolic equivalent of task

**OFC**: orbitofrontal cortex

**VLPFC**: ventrolateral prefrontal cortex

**VR**: virtual reality

**ΔacHbO2**: change in accumulated oxygenated hemoglobin

---

Edited by N Zary, G Eysenbach; submitted 20.06.22; peer-reviewed by J O'Neil, S Pandey; comments to author 04.08.22; revised version received 28.09.22; accepted 31.10.22; published 05.01.23.

Please cite as:

Kim HJ, Lee JW, Choi G, Huh J, Han DH

Differences in Brain Activity and Body Movements Between Virtual Reality and Offline Exercise: Randomized Crossover Trial

JMIR Serious Games 2023;11:e40421

URL: https://games.jmir.org/2023/1/e40421

doi:10.2196/40421

PMID:36602842

©Hee Jin Kim, Jea Woog Lee, Gangta Choi, Junghoon Huh, Doug Hyun Han. Originally published in JMIR Serious Games (https://games.jmir.org), 05.01.2023. 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.
Measured and Perceived Exercise Intensity During the Performance of Single-Task, Cognitive-Motor Dual-Task, and Exergame Training: Transversal Study

Matthieu Gallou-Guyot¹, PT, PhD; Anaïck Perrochon¹, PhD; Romain Marie², PhD; Maxence Bourgeois¹, PT; Stéphane Mandigout¹, PhD

¹HAVAE, UR 20217, Université de Limoges, Limoges, France
²3iL Ingénieurs, Limoges, France

Abstract

Background: The physical and cognitive loads borne during exergaming may differ from more conventional cognitive-motor dual-task trainings.

Objective: The aim of this pilot transversal study was to compare objectively measured and perceived exercise intensity during exergame, cognitive-motor dual-task, and single-task training sessions.

Methods: We recruited apparently healthy young adults who carried out one session of each type of training: exergaming, cognitive-motor dual-tasking, and single-tasking. We used a custom-made exergame as support. The sessions lasted 30 minutes, were spaced at least 24 hours apart, and took place in random order for each group of 4 participants. We used heart rates to assess exercise intensity and the modified Borg scale to assess perception of intensity. In all, 16 apparently healthy young participants carried out all sessions.

Results: There was no difference between the different types of training in mean heart rates (P=.27), peak heart rates (P=.50), or Borg scale scores (P=.40). Our custom-made exergame’s objectively measured and perceived physical load did not differ between cognitive-motor dual-task and single-task training.

Conclusions: As a result, our exergame can be considered to be as challenging as more traditional physical training. Future studies should be conducted in older adults with or without cognitive impairments and incorporate an assessment of cognitive performance.

(JMIR Serious Games 2023;11:e36126) doi:10.2196/36126

KEYWORDS

exergame; dual-task; exercise intensity; heart rate; cognitive load; active video game; physical activity

Introduction

Exergames (EGs), or “active video games,” are video games played on a digital device, including a wide range of interfaces [1] that require physical activity when played [2]. EGs are a singular form of cognitive-motor dual-task (CMDT) training, which is known to be efficient in terms of both cognitive and physical functions [3,4]. In addition, EGs are considered fun and are enjoyed by most users, as they facilitate exercise in an attractive, motivating, and interactive environment [5]. However, EGs must be sufficiently demanding to induce effects, with regard to cognitive as well as physical loads [2,6], and their efficiency depends to some extent on their intensity.

As a singular form of CMDT training, EGs should not differ in terms of the physical level of solicitation. The tasks players are asked to perform are similar, one of them being cognitive and...
the other physical, with the difference residing in the support used. Previous studies have shown that most EGs induce moderate [7-10] to vigorous [8,9,11] exercise intensity in healthy young adults. However, there seem to be large variations in intensity between different EGs, depending on the hardware (primarily Kinect or Wii) or game used [11,12]. For example, EGs such as Wii Bowling induce low-intensity physical activity [8], whereas Kinect Boxing or “Your Shape Fitness Evolved 2012” indeed induce vigorous physical activity [9,11]. Variability in exercise intensity between tailor-made EGs is likely to be even greater.

Significant heterogeneity in exercise intensity during exergaming raises questions about the impact of the support used. Would exercise intensity be the same during similar physical training sessions presented in either single-task (ST) or CMDT conditions, whether using a tailor-made EG as a support or not? To our knowledge, no study has compared exercise intensity during exergaming and ST training, and only one study has compared exercise intensity during CMDT and ST training in healthy young adults [13], showing higher intensity under CMDT than ST conditions.

To date, we have been unable to conclusively determine the impact of the training support (ie, direct comparison between exergaming and CMDT training) or the impact of a concurrent cognitive task (ie, direct comparison to ST conditions, considered as the reference modality for physical training) on the level of physical solicitation in healthy young adults. The aim of this pilot transversal study was to compare objectively measured and perceived exercise intensity during EG, CMDT, and ST training sessions.

**Methods**

**Participants**

Apparently healthy students from the University of Limoges volunteered for this pilot transversal study. The inclusion criteria were young adults aged between 18 and 35 years and fluent in French. The exclusion criteria were having a contraindication to physical activity or having eaten or drunk during the previous 2 hours.

**Ethical Considerations**

In accordance with the Declaration of Helsinki as revised in 2013, the volunteers had received an information document detailing the protocol and gave their written consent to participate.

The study was carried out as part of student work in initiation to research during the year 2021. According to French regulations (Jardé law) [14], ethics review was not required.

**Procedure**

Participants were recruited during practical exercise courses. They carried out 3 types of training: EG, CMDT, and motor ST. The training sessions were designed identically: physical exercises (ST), with concurrent cognitive tasks (CMDT), using an EG as support (EG). The exercise sessions were organized in groups of 4 participants, and the order of training was randomized for each group. They lasted 30 minutes and were spaced at least 24 hours apart. They were divided into eight 3-minute-long exercise sequences (see Table 1), with 30 seconds of rest and instructions in between; therefore, effective physical activity lasted approximately 24 minutes. An instructor supervised, giving instructions and assisting the participants. Sessions were suspended in the event of an injury, significant pain, or heart rates above 90% of the theoretical maximums [15]. We have illustrated the differences between EG, CMDT, and ST training during stepping performance in Figure 1.

During EG training, sessions consisted of multiple dual tasks using a custom-designed EG as support (Figure 2). This EG was developed in the Handicap, Activités Vieillissement, Autonomie, Environnement laboratory, using the “Virtual Carpet,” which is associated a video projector and HTC Vive cameras and trackers, as the play area [16-18]. The projected scene was a schematic city, and players were asked to move to different points of interest during exercises. Detection by the Vive trackers of a player’s position at a point of interest triggered a change of scene and launched a mini-game, which was carried out by all the players. The motor tasks were mainly stepper, muscle strength, and balance activities. The cognitive tasks were verbal fluency, arithmetic, mental inhibition and flexibility, visuospatial memory, processing speed, and planning. Some exercises required gymnastic equipment (stepper, Swiss or medicine balls, chairs). The training sessions were designed to match with the American College of Sports Medicine [19] and the World Health Organization [20] guidelines and recommendations on physical activity, as well as the recommendations designed to prevent falls in older adults [21]. Details on the exercises are indicated in Table 1.

During the CMDT training, the participants used the same gymnastic equipment, and the sessions were made up of the same associations of cognitive and motor dual tasks as in the EG group but without using the EG (see exercise details in Table 1).

Finally, ST training required the same gymnastic equipment and the same motor tasks as those carried out during EG and CMDT training but without a concurrent cognitive task (see exercise details in Table 1).
### Table 1. Details of the exercises proposed during the different types of training.

<table>
<thead>
<tr>
<th>Training</th>
<th>Stepper (with and without a step)</th>
<th>Visuospatial memory and balance</th>
<th>Muscular strength and coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EG</strong></td>
<td>Arrows are displayed successively on the projected scene, and participants must reproduce them on a pad with 1 foot, 2 feet, doing a squat, a lunge, etc. The additional cognitive tasks are to not reproduce an arrow (go or no go), to invert them (mental flexibility), or to render them with a time lapse (working memory).</td>
<td>Eight elements displayed within the projected area will turn on and off at a fixed frequency, constituting a growing span. Participants must memorize this sequence and then recall it while moving around. At the same time, they perform motor exercises (knee raising, buttocks to heels, squats, lunges, and jumping jacks). The additional cognitive tasks are to not consider one of the icons or to recall the span from the end.</td>
<td>Participants must perform muscle-strengthening exercises (eg, squat and lunges). At the same time, they must (1) solve mental arithmetic exercises appearing on the projected scene; (2) alternate the exercises performed according to the images appearing on the projected scene; and (3) perform a “categories” game or build a word giving a letter, one at a time.</td>
</tr>
<tr>
<td><strong>CMDT</strong></td>
<td>The instructor shows a sequence of movements that the participants must reproduce in mirror (step, squat, lunge, etc). The additional cognitive tasks are to not reproduce an arrow (go or no go), to invert them (mental flexibility), or to render them with a time lapse (working memory).</td>
<td>Participants must displace a total of 8 plots, one at a time. The plot-placing order constitutes the span. Participants must memorize this sequence and then recall it while moving around. At the same time, they perform motor exercises (knee raising, buttocks to heels, squats, lunges, and jumping jacks). The additional cognitive tasks are to not consider one of the icons or to recall the span from the end.</td>
<td>Participants must perform muscle-strengthening exercises (ie, squats and lunges). At the same time, they must (1) alternate the exercises performed according to the auditory or visual stimuli given by the instructor and (2) perform a “categories” game or build a word giving a letter, one at a time.</td>
</tr>
<tr>
<td><strong>ST</strong></td>
<td>The instructor shows a sequence of movements that the participants must reproduce in mirror (step, squat, lunge, etc). There is no additional cognitive task.</td>
<td>The instructor demonstrates static and dynamic balance exercises to perform. The participants must perform movements of the limbs and trunk and pass a medicine ball while standing on a Swiss ball or standing on one leg. There is no additional cognitive task.</td>
<td>Participants perform a game mixing together muscle strengthening of the lower limbs and motor coordination by dribbling with a ball. There is no additional cognitive task.</td>
</tr>
</tbody>
</table>

---

**a**EG: exergame.

**b**CMDT: cognitive-motor dual-task.

**c**ST: single-task.

Figure 1. Illustration of the differences between exergame (EG), cognitive-motor dual-task (CMDT), and single-task (ST) training, during the performance of a stepping task. The instructor is represented in blue.
Figure 2. Illustration of the exergame. (1) Video projector, (2) HTC Vive infrared cameras, (3) computer, and (4) HTC Vive trackers. In this example, only 2 participants are performing a stepping exercise.

Outcomes

The primary outcome was the objectively measured exercise-intensity level assessed in terms of mean and peak heart rates (HRs) during training. We used HR because it is considered the most practical parameter to monitor, especially as regards reliability, safety, and cost [22]. We used validated chest HR Polar H10 monitors (Polar Electro Oy) [23], measuring the participants’ $H_{\text{mean}}$ and $H_{\text{peak}}$ during the 3 types of training.

The secondary outcome was the rating of perceived exertion, using a validated modified Borg scale [24] completed by participants at the end of each training session.

Statistical Analysis

Quantitative variables were described according to mean and SD or median and IQR. The normality of data distribution was assessed using Shapiro-Wilk tests. Comparison between $H_{\text{mean}}$, $H_{\text{peak}}$, and Borg scale scores during the different trainings (EG, CMDT, and ST) was carried out using a 1-factor ANOVA or a Friedman test according to the normality of the variables concerned. Statistical analysis was performed using RStudio (Rstudio, Inc), and result significance was set at $P<.05$.

Results

In all, 16 apparently healthy young participants volunteered for the study (6 males, 10 females). Sociodemographic characteristics are presented in Table 2. Although mean and peak HR variable distribution was normal during the 3 types of training, the Borg scale scores were not. There was no difference between the different types of training (EG, CMDT, and ST) for any of the variables studied: $H_{\text{mean}}$ ($F_{2,45}=1.33; P=.27$), $H_{\text{peak}}$ ($F_{2,45}=0.7; P=.50$), and Borg scale scores ($\chi^2=1.85, P=.40$; see details in Table 3).

Table 2. Participant characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>24.6 (3.1)</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.72 (0.11)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>69.1 (15.1)</td>
</tr>
<tr>
<td>BMI</td>
<td>23.1 (3.2)</td>
</tr>
<tr>
<td>$H_{\text{resting}}$ (bpm)$^b$</td>
<td>66.9 (10.4)</td>
</tr>
<tr>
<td>$H_{\text{max}}$ (bpm)</td>
<td>187.2 (1.2)</td>
</tr>
</tbody>
</table>

$^a$HR: heart rate.

$^b$bpm: beats per minute.
Previous studies assessing the level of solicitation during exergaming in healthy young adults through HR have shown moderate [7-10] or vigorous [8,9,11] exercise intensity, with highly pronounced variability between studies according to the software or hardware used. Our custom-made EG seemed to induce moderate exercise intensity, equivalent to the physical exercises requested under CMDT or ST conditions. This is a relevant and interesting finding; since exercises under ST conditions are considered as the reference modality for physical training, their equivalence during exergaming and dual-task training enabled us to consider our EG as sufficiently soliciting to induce physical results, while adding a cognitive load in a playful environment.

Moreover, perceived physical intensity assessed through the Borg scale did not differ between the 3 types of training and was globally moderate. The perception of effort by participants in this study was correlated with the results obtained from the theoretical HR\textsubscript{max} [25]. This is in line with previous studies showing that adults correctly estimate their percentage of theoretical HR\textsubscript{max}, thereby limiting impact of interpersonal variability. This pilot transversal study was aimed at comparing objectively measured and perceived exercise intensity during EG, CMDT, and ST training sessions in healthy young adults. We observed no differences in mean or peak HR or perceived exertion between the 3 types of training.

### Perceived and Measured Physical Activity Intensity

In our study, physical activity intensities during EG, CMDT, and ST training sessions did not statistically differ. This finding is discrepant with the results of a previous study, which showed higher HR\textsubscript{mean} under CMDT than ST training [13]. Our results even seemed opposed, with a tendency toward higher HR\textsubscript{mean} during ST training of nearly 10 beats per minute, close to 70% of theoretical HR\textsubscript{max}, and moderate intensity [25]. This difference with our study may be due to the nature of the requested tasks. Exercise intensity varies considerably according to type of movement, the body part involved (upper or lower limb), the level of difficulty of a given game (frequency and speed), or the participants’ previous experience as players [26-28]. In the study by Herold et al [13], subjects were asked to carry out squats under both conditions while additionally counting backward in CMDT training [13]. In our study, the physical and cognitive tasks were not only more varied but also and above all more complex (see Table 1). In addition, it is important to note the impact that different interfaces can have when comparing interventions. For instance, the quality of immersion and interaction highly impacts the user experience and the usability of immersive EGs [29]. We tried to reduce these effects of complexity and interface by proposing similar and transferable exercises between the different types of training. As a result, the exercises carried out in this study during training sessions were the same in EG, CMDT and ST conditions, with the differences consisting solely of the concurrent cognitive task and the use of a given game as support. In fact, instruction comprehension and error correction were time-consuming, decreasing effective exercise duration and HR over the course of the 30-minute sessions. Understanding more complex combined tasks during CMDT training along with adaptation to new game rules and to a new environment during EG may have been more time-consuming than during ST training.

### Discussion

This pilot transversal study was aimed at comparing objectively measured and perceived exercise intensity during EG, CMDT, and ST training sessions in healthy young adults. We observed no differences in mean or peak HR or perceived exertion between the 3 types of training.

#### Perceived and Measured Physical Activity Intensity

In our study, physical activity intensities during EG, CMDT, and ST training sessions did not statistically differ. This finding is discrepant with the results of a previous study, which showed higher HR\textsubscript{mean} under CMDT than ST training [13]. Our results even seemed opposed, with a tendency toward higher HR\textsubscript{mean} during ST training of nearly 10 beats per minute, close to 70% of theoretical HR\textsubscript{max}, and moderate intensity [25]. This difference with our study may be due to the nature of the requested tasks. Exercise intensity varies considerably according to type of movement, the body part involved (upper or lower limb), the level of difficulty of a given game (frequency and speed), or the participants’ previous experience as players [26-28]. In the study by Herold et al [13], subjects were asked to carry out squats under both conditions while additionally counting backward in CMDT training [13]. In our study, the physical and cognitive tasks were not only more varied but also and above all more complex (see Table 1). In addition, it is important to note the impact that different interfaces can have when comparing interventions. For instance, the quality of immersion and interaction highly impacts the user experience and the usability of immersive EGs [29]. We tried to reduce these effects of complexity and interface by proposing similar and transferable exercises between the different types of training. As a result, the exercises carried out in this study during training sessions were the same in EG, CMDT and ST conditions, with the differences consisting solely of the concurrent cognitive task and the use of a given game as support. In fact, instruction comprehension and error correction were time-consuming, decreasing effective exercise duration and HR over the course of the 30-minute sessions. Understanding more complex combined tasks during CMDT training along with adaptation to new game rules and to a new environment during EG may have been more time-consuming than during ST training.

#### Limitations

The first limitation of this study is the sample, for which no a priori sample size was carried out. However, our population was homogeneous and showed relative stability in physical intensity. This pilot study constitutes a proof of concept specific to our EG in a highly specific context and population. Consequently, the small extension of the knowledge provided by our study is specific, which complicates its extrapolation to other forms of EGs or populations.

A second limitation is due to the assessment exercise intensity modalities. In this study, given its reliability, safety, and cost, we used HR as a reflection of exercise intensity, whereas other studies have used more direct parameters such as oxygen consumption [8,31] or blood lactate responses [32] to assess exercise intensity during exergaming. Moreover, the software we used did not enable us to analyze HR in detail. It would be relevant to estimate the HR profile to determine whether it remains stable and close to the mean or varies with peaks and rest. Lastly, the assessment of effective HR\textsubscript{max} rather than theoretical HR\textsubscript{max} would have made it possible to evaluate reserve HR, thereby limiting impact of interpersonal variability.

#### Table 3. Results depending on type of training.

<table>
<thead>
<tr>
<th>Result</th>
<th>EG\textsuperscript{a}</th>
<th>CMDT\textsuperscript{b}</th>
<th>ST\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borg scale score (out of 10), mean (SD)</td>
<td>3.1 (1.4)</td>
<td>3.6 (1.3)</td>
<td>3.3 (1.0)</td>
</tr>
<tr>
<td>HR\textsubscript{mean} (bpm\textsuperscript{d}), mean (SD)</td>
<td>119.8 (12.2)</td>
<td>123.8 (16.1)</td>
<td>128.1 (14.7)</td>
</tr>
<tr>
<td>Percentage of theoretical HR\textsubscript{max} (%)</td>
<td>64</td>
<td>66</td>
<td>68</td>
</tr>
<tr>
<td>HR\textsubscript{peak} (bpm), mean (SD)</td>
<td>157.9 (10.1)</td>
<td>163.0 (12.8)</td>
<td>161.7 (14.5)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}EG: exergame.
\textsuperscript{b}CMDT: cognitive-motor dual-task.
\textsuperscript{c}ST: single-task.
\textsuperscript{d}HR: heart rate.
\textsuperscript{e}bpm: beats per minute

Perceived and Measured Physical Activity Intensity

In our study, physical activity intensities during EG, CMDT, and ST training sessions did not statistically differ. This finding is discrepant with the results of a previous study, which showed higher HR\textsubscript{mean} under CMDT than ST training [13]. Our results even seemed opposed, with a tendency toward higher HR\textsubscript{mean} during ST training of nearly 10 beats per minute, close to 70% of theoretical HR\textsubscript{max}, and moderate intensity [25]. This difference with our study may be due to the nature of the requested tasks. Exercise intensity varies considerably according to type of movement, the body part involved (upper or lower limb), the level of difficulty of a given game (frequency and speed), or the participants’ previous experience as players [26-28]. In the study by Herold et al [13], subjects were asked to carry out squats under both conditions while additionally counting backward in CMDT training [13]. In our study, the physical and cognitive tasks were not only more varied but also and above all more complex (see Table 1). In addition, it is important to note the impact that different interfaces can have when comparing interventions. For instance, the quality of immersion and interaction highly impacts the user experience and the usability of immersive EGs [29]. We tried to reduce these effects of complexity and interface by proposing similar and transferable exercises between the different types of training. As a result, the exercises carried out in this study during training sessions were the same in EG, CMDT and ST conditions, with the differences consisting solely of the concurrent cognitive task and the use of a given game as support. In fact, instruction comprehension and error correction were time-consuming, decreasing effective exercise duration and HR over the course of the 30-minute sessions. Understanding more complex combined tasks during CMDT training along with adaptation to new game rules and to a new environment during EG may have been more time-consuming than during ST training.

**Previous studies assessing the level of solicitation during exergaming in healthy young adults through HR have shown moderate [7-10] or vigorous [8,9,11] exercise intensity, with highly pronounced variability between studies according to the software or hardware used. Our custom-made EG seemed to induce moderate exercise intensity, equivalent to the physical exercises requested under CMDT or ST conditions. This is a relevant and interesting finding; since exercises under ST conditions are considered as the reference modality for physical training, their equivalence during exergaming and dual-task training enabled us to consider our EG as sufficiently soliciting to induce physical results, while adding a cognitive load in a playful environment.**
Future Studies

Most of the EGs in which the intensity was assessed in previous studies were commercial video games [7-11]. This is broadly the case, especially insofar as commercial active video games are frequently used in rehabilitation [33]. The main consequence is a lack of control over qualitative and quantitative physical and cognitive exercises. Tailor-made software using mainstream hardware is designed to suit the specific needs of a given audience [34] while applying an easily available and relatively inexpensive solution. For example, the training of older adults presents specificities, in both form and in substance. Understanding is maximized through slow animations, large fronts, and simple rules [35]. With age, physical and cognitive capacities decline while cognitive-motor interference increases [36,37]. Fall prevention in older adults should focus on strength, postural control, stepping, and gait training [38-40], along with mental inhibition and flexibility, processing speed, and visuospatial memory [41-43]. In fact, a fall prevention training protocol with our EG seems possible and relevant, insofar as it facilitates the simultaneous training of cognitive and physical functions. By applying our pilot study’s values and methodology, future studies may assess the extent to which and the efficacy with which our EG solicits in older adults. In addition to its pronounced beneficial effects, our EG may help promote physical activity [44], which is a point of major importance among older adults.

In this study, we investigated physical aspects, not cognition. However, cognitive efficiency and HR may be correlated; a meta-analysis found that participants with better cognitive task results often demonstrated greater cardiac parasympathetic control than those with poorer cognitive performance [45]. This question could be investigated using hybrid systems, coupling functional near infrared spectroscopy, and electroencephalogram headsets [46].

Conclusion

This pilot transversal study showed that a custom-made EG that could induce moderate perceived and objectively measured exercise intensity, equivalent to CMDT and ST training in healthy young adults. As a result, our EG can be considered as being as relevant as more traditional physical training with regard to exercise intensity. Future studies should investigate the cognitive and physical level of solicitation of our EG in older adults, who are likely to draw benefit from CMDT training.

Acknowledgments

The authors thank all the participants who volunteered, as well as the City of Limoges for their support. The authors also thank Dr Fanny Thomas for her contribution.

This research received a grant from the City of Limoges and the Nouvelle Aquitaine region. The funding source had no involvement in the conduct of the research.

Authors’ Contributions

MG-G contributed to conceptualization, methodology, investigation, writing–original draft, and visualization. AP contributed to conceptualization, methodology, writing–original draft, and project administration. RM contributed to development and writing–review and editing. MB contributed to investigation. SM contributed to conceptualization, methodology, writing–original draft, project administration, and supervision.

Conflicts of Interest

None declared.

References


https://games.jmir.org/2023/1/e36126 JMIR Serious Games 2023 | vol. 11 | e36126 | p.316 (page number not for citation purposes)


Abbreviations

CMDT: cognitive-motor dual-task
EG: exergame
HR: heart rate
ST: single-task
Please cite as:
Gallou-Guyot M, Perrochon A, Marie R, Bourgeois M, Mandigout S
Measured and Perceived Exercise Intensity During the Performance of Single-Task, Cognitive-Motor Dual-Task, and Exergame Training: Transversal Study
JMIR Serious Games 2023;11:e36126
URL: https://games.jmir.org/2023/1/e36126
doi:10.2196/36126
PMID:36729572
Research Status and Emerging Trends in Virtual Reality Rehabilitation: Bibliometric and Knowledge Graph Study

Ting Fan¹, MD; Xiaobei Wang², BNAD; Xiaoxi Song³, MD; Gang Zhao⁴, PhD; Zhichang Zhang¹, MEng
¹Department of Computer, School of Intelligent Medicine, China Medical University, Shenyang, China
²Department of General Practice, The First Hospital of China Medical University, China Medical University, Shenyang, China
³Liaoning Education Informatization Construction Center, LiaoNing Institute of Education, Shenyang, China
⁴Department of Health Promotion, School of Intelligent Medicine, China Medical University, Shenyang, China

Corresponding Author:
Zhichang Zhang, MEng
Department of Computer
School of Intelligent Medicine
China Medical University
No 77 Puhe Road
Shenyang North New Area
Shenyang, 110122
China
Phone: 86 18900910770
Email: zczhang@cmu.edu.cn

Abstract

Background: Virtual reality (VR) technology has been widely used in rehabilitation training because of its immersive, interactive, and imaginative features. A comprehensive bibliometric review is required to help researchers focus on future directions based on the new definitions of VR technologies in rehabilitation, which reveal new situations and requirements.

Objective: Herein, we aimed to summarize effective research methods for and potential innovative approaches to VR rehabilitation by evaluating publications from various countries to encourage research on efficient strategies to improve VR rehabilitation.

Methods: The SCIE (Science Citation Index Expanded) database was searched on January 20, 2022, for publications related to the application of VR technology in rehabilitation research. We found 1617 papers, and we created a clustered network, using the 46,116 references cited in the papers. CiteSpace V (Drexel University) and VOSviewer (Leiden University) were used to identify countries, institutions, journals, keywords, cocited references, and research hot spots.

Results: A total of 63 countries and 1921 institutes have contributed publications. The United States of America has taken the leading position in this field; it has the highest number of publications; the highest h-index; and the largest collaborative network, which includes other countries. The reference clusters of SCIE papers were divided into the following nine categories: kinematics, neurorehabilitation, brain injury, exergames, aging, motor rehabilitation, mobility, cerebral palsy, and exercise intensity. The research frontiers were represented by the following keywords: video games (2017-2021), and young adults (2018-2021).

Conclusions: Our study comprehensively assesses the current research state of VR rehabilitation and analyzes the current research hot spots and future trends in the field, with the aims of providing resources for more intensive investigation and encouraging more researchers to further develop VR rehabilitation.

(JMIR Serious Games 2023;11:e41091) doi:10.2196/41091

KEYWORDS
mobility; rehabilitation; virtual reality; bibliometric; technology; training; interactive; research; exercise; resources; cerebral palsy; adult; video games

Introduction

In recent years, the number of people with rehabilitation needs has increased, particularly among groups of older patients, patients with disabilities, patients with chronic diseases, and patients with functional and cognitive impairments. The loss of movement, sensation, balance, and cognition, as well as other aspects, seriously affects patients’ quality of life, work, study,
and social life [1,2]. Such patients require long-term, consistent rehabilitation training and guidance [3]. However, traditional rehabilitation training has a number of problems, including fixed rehabilitation centers, a lack of rehabilitation resources, uninteresting training processes, high treatment costs, and a lack of automatic guidance and incentive mechanisms. These result in a lack of confidence in the rehabilitation process, which in turn affects the outcomes of rehabilitation treatments [4,5].

With the gradual popularization of virtual reality (VR) technology, rehabilitation training systems based on VR technology have been gradually applied in sports, exercise, and functional rehabilitation for various diseases and have achieved positive effect results [6,7]. The combination of VR technology and rehabilitation medicine can enable more patients to train regularly at home or in the community, as VR rehabilitation systems provide an immersive experience that stimulates patients’ interest and improves their participation, thus overcoming the disadvantages of fixed centers and the lack of resources [8]. Furthermore, VR rehabilitation systems can sense and record a patient’s movement and biological data via sensors to further improve existing rehabilitation programs [9]. This rehabilitation technology is a useful supplement to traditional rehabilitation and is a promising new research direction in the field of rehabilitation medicine.

A comprehensive bibliometric review is required to help researchers focus on future directions based on the new definitions of VR technologies in rehabilitation, which reveal new situations and requirements. Although bibliometric methods have yielded positive results in a variety of fields, we found that there is still a significant gap in the research on VR rehabilitation and its development trends by using bibliometric methods.

We used bibliometric methods to analyze SCIE (Science Citation Index Expanded) papers on studies related to VR rehabilitation research. Articles from different countries, regions, and research institutions were included. We identified papers in journals, gathered the top 10 citations, and enumerated how many times these citations were used. The VR rehabilitation knowledge base was analyzed by grouping authors’ co-occurring keyword networks. Burst citations were used to identify research hot spots on this topic, which could provide a useful reference for future research [10,11]. These analyses will provide rehabilitation specialists with a macroscopic understanding of the knowledge domain as a whole, as well as a microscopic characterization. Compared to other reviews, our study is timely and visual and provides an impartial approach to developing and exploring particular knowledge domains. Our findings may encourage more researchers to conduct additional research in this field to further develop VR rehabilitation methods. The following basic information was gathered from studies: titles, abstracts, author information, institutions, countries, regions, keywords, and citations.

### Methods

We created a clustered network, using the 46,116 references cited in 1617 published papers. The data came from the Web of Science Core Collection (WSCC), which was accessed on January 20, 2022. The search covered articles published from 2000 to 2021. The following search string, which included the search terms, was used: 

\[
\text{Virtual Reality OR VR OR AR OR Augmented Reality}
\]

Further, we limited our search to original articles and reviews, and the Web of Science category Rehabilitation was selected. To determine study inclusion, the aforementioned basic information was collected from articles in the WSCC. However, the following papers were excluded: (1) irrelevant proceedings and meetings papers, (2) chapters in books, (3) duplicated articles, and (4) unpublished papers with limited information. In total, 1617 papers were included, with duplicates excluded. The search and analysis procedures are outlined in Figure 1.

We defined most publication traits, including institute, country, journal, and keywords. We used Journal Citation Reports (2020 version; Clarivate) to identify impact factors, which reflect scientific research value [12]. All data were processed by using CiteSpace V (Drexel University) and VOSviewer (Leiden University). Both programs are used for collaboration network analysis to link publication traits [11,13]. From these analyses and measures, we obtained data on research hot spots, evolution paths, knowledge structures, and new trends in VR rehabilitation.
Results

Study Distribution by Publication Year

Between 2000 and 2021, a total of 1617 papers on VR rehabilitation were published. Emerging trends in VR rehabilitation research–related studies are outlined in Figure 2. From 2000 onward, VR rehabilitation research rapidly accelerated. In Figure 2, the blue line denotes the increasing trend in the annual number of studies from 2000 to 2021, whereas the red broken line represents the publication index, which also increased. In the initial 2000 to 2006 research period, publication numbers per year were relatively stable and were <25, suggesting an initial period of exploratory VR rehabilitation research. However, the rise in rehabilitation demands and the increased development of supporting technologies, which are due to the increased interest in applying VR to medicine and rehabilitation and the increasing, widespread use of VR technology, have resulted in a proliferation of publications since then. The number of publications in last decade accounted for more than 80% of the total publications found.

Analysis of Countries and Institutes

The analyzed publications included contributions from a total of 63 countries and 1921 institutes. The top 10 countries and institutes are outlined in Tables 1 and 2, respectively. The United States published the most studies (n=663), followed by Canada (n=143). The United States also had the top h-index (59), followed by Canada (29) and Australia (21). Collaborations between countries and between institutes are shown in Figure 3 and Figure 4, respectively.
University New Brunswick had the most publications (n=52), followed by McGill University (n=45). Rutgers State University New Brunswick also had the highest h-index (27). On a global level, research institutes and associated research staff are collaborating and sharing experiences. However, although the United States has played a leading role, the international collaboration rates of the United States are low.

Table 1. The top 10 countries in terms of publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Countries</th>
<th>Publication count, n</th>
<th>h-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States of America</td>
<td>663</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>Canada</td>
<td>143</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>Australia</td>
<td>96</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>England</td>
<td>92</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>South Korea</td>
<td>83</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>Italy</td>
<td>81</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Israel</td>
<td>71</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>Brazil</td>
<td>68</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>Netherlands</td>
<td>67</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>Spain</td>
<td>61</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 2. Top 10 institutions in terms of publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institutions</th>
<th>Publication count, n</th>
<th>h-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rutgers State University New Brunswick</td>
<td>52</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>McGill University</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Pennsylvania Commonwealth System of Higher Education</td>
<td>38</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Tel Aviv University</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>University of Wisconsin System</td>
<td>38</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>University of Haifa</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>University of Illinois System</td>
<td>35</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>US Department of Veterans Affairs</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Veterans Health Administration</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>University of Wisconsin Madison</td>
<td>33</td>
<td>11</td>
</tr>
</tbody>
</table>
Figure 3. The cooperation of countries and regions that contributed to publications on virtual reality rehabilitation. CC: co-citations; CST: Central Standard Time; LBY: Look back year; L/N: Maximum Links Per Node; LRF: Link Retaining Factor; WoS: Web of Science.
Analysis of Journals

A total of 241 journals published articles related to VR rehabilitation; the top 10 are outlined in Table 3, and collaborations between cited journals are shown in Figure 5. Archives of Physical Medicine and Rehabilitation was the top journal (publications: n=847), followed by Physical Therapy and Rehabilitation (publications: n=608). Among these journals, Cochrane Database of Systematic Reviews had the highest impact factor (9.266), followed by Stroke (7.914).

Table 3. The top 10 journals in terms of publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Cited journals</th>
<th>Publication count, n</th>
<th>h-index</th>
<th>Impact factor (2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Archives of Physical Medicine and Rehabilitation</td>
<td>847</td>
<td>197</td>
<td>3.966</td>
</tr>
<tr>
<td>2</td>
<td>Physical Therapy and Rehabilitation</td>
<td>608</td>
<td>150</td>
<td>3.021</td>
</tr>
<tr>
<td>3</td>
<td>Disability and Rehabilitation</td>
<td>550</td>
<td>111</td>
<td>3.033</td>
</tr>
<tr>
<td>4</td>
<td>Journal of NeuroEngineering and Rehabilitation</td>
<td>550</td>
<td>94</td>
<td>4.262</td>
</tr>
<tr>
<td>5</td>
<td>Stroke</td>
<td>484</td>
<td>319</td>
<td>7.914</td>
</tr>
<tr>
<td>6</td>
<td>Neurorehabilitation and Neural Repair</td>
<td>454</td>
<td>106</td>
<td>3.919</td>
</tr>
<tr>
<td>7</td>
<td>CyberPsychology &amp; Behavior</td>
<td>420</td>
<td>143</td>
<td>4.157</td>
</tr>
<tr>
<td>8</td>
<td>Clinical Rehabilitation</td>
<td>411</td>
<td>110</td>
<td>3.477</td>
</tr>
<tr>
<td>9</td>
<td>Journal of Rehabilitation Medicine</td>
<td>357</td>
<td>96</td>
<td>2.912</td>
</tr>
<tr>
<td>10</td>
<td>Cochrane Database of Systematic Reviews</td>
<td>324</td>
<td>273</td>
<td>9.266</td>
</tr>
</tbody>
</table>
Analyzing Citations

A citation is a vital bibliometric indicator, with frequently cited studies greatly influencing their research areas. We list the top 10 most highly cited publications in Table 4. *Effectiveness of Virtual Reality Using Wii Gaming Technology in Stroke Rehabilitation* by Saposnik et al [14], which was published on 2010 in *Stroke*, was the most highly cited article (cited 58 times); the article verified the effectiveness of rehabilitation for participants with stroke by using Wii (Nintendo Co, Ltd) gaming technology. Laver et al [15,16] published 2 articles with the same title—*Virtual reality for stroke rehabilitation*—on 2011 [15] and 2015 [16] in *Cochrane Database of Systematic Reviews*. These studies evaluated the rehabilitation impact of
VR technology for patients with stroke in accordance with the status quo in different periods.

In network research, betweenness centrality reflects node importance in a network; thus, a higher betweenness centrality signifies an important study [23]. The betweenness centralities of the top 10 studies are shown in Table 4.

A cocited document–centered clustering investigation was proposed to identify subnodes and connecting nodes in VR rehabilitation. To assess the scientific relevance of publications, we established a network of cocited references (Figure 7). The cluster settings were as follows: the number of years per slice was set to 1, and the top 0.5% of articles were selected for analysis; a pruning algorithm was used. The modularity Q score was 0.8029 (>0.5), which showed that the network was well separated into loosely coupled clusters. The weighted mean silhouette score was 0.941 (>0.5) and suggested acceptable cluster homogeneity. Study index items were used as cluster markers. The largest cluster—cluster 0—was labeled as *kinematics*, cluster 1 was labeled as *neurorehabilitation*, cluster 2 was labeled as *brain injury*, cluster 3 was labeled as *exergame*, cluster 4 was labeled as *aging*, cluster 5 was labeled as *motor rehabilitation*, cluster 6 was labeled as *mobility*, cluster 7 was labeled as *cerebral palsy*, and cluster 8 was labeled as *exercise intensity*.

Table 4. The top 10 cited articles on virtual reality (VR) in rehabilitation.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Authors and published year</th>
<th>Title of cited article</th>
<th>Citation count, n</th>
<th>Centrality</th>
<th>Interpretation of the findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Saposnik et al [14], 2010</td>
<td>Effectiveness of Virtual Reality Using Wii Gaming Technology in Stroke Rehabilitation</td>
<td>58</td>
<td>0.04</td>
<td>This article verified the effectiveness of rehabilitation for participants with stroke by using Wii (Nintendo Co, Ltd) gaming technology.</td>
</tr>
<tr>
<td>2</td>
<td>Saposnik and Levin [17], 2011</td>
<td>Virtual Reality in Stroke Rehabilitation: A Meta-Analysis and Implications for Clinicians</td>
<td>51</td>
<td>0.10</td>
<td>This study performed a meta-analysis to determine the added benefit of VR technology in arm motor recovery after stroke.</td>
</tr>
<tr>
<td>3</td>
<td>Deutsch et al [18], 2008</td>
<td>Use of a Low-Cost, Commercially Available Gaming Console (Wii) for Rehabilitation of an Adolescent With Cerebral Palsy</td>
<td>37</td>
<td>0.04</td>
<td>This document is the first report on using low-cost, commercially available gaming technology for the rehabilitation of an adolescent with cerebral palsy.</td>
</tr>
<tr>
<td>4</td>
<td>Laver et al [15], 2011</td>
<td>Virtual reality for stroke rehabilitation</td>
<td>37</td>
<td>0.12</td>
<td>This study evaluated the effects of VR and interactive video gaming on upper limb, lower limb, and global motor function.</td>
</tr>
<tr>
<td>5</td>
<td>Lohse et al [6], 2014</td>
<td>Virtual Reality Therapy for Adults Post-Stroke: A Systematic Review and Meta-Analysis Exploring Virtual Environments and Commercial Games in Therapy</td>
<td>31</td>
<td>0.14</td>
<td>This analysis systematically reviewed the evidence for VR therapy in adults after stroke.</td>
</tr>
<tr>
<td>6</td>
<td>Levin et al [19], 2015</td>
<td>Emergence of Virtual Reality as a Tool for Upper Limb Rehabilitation: Incorporation of Motor Control and Motor Learning Principles</td>
<td>27</td>
<td>0.01</td>
<td>This article discussed how to exploit VR training environments and provided evidence concerning applications for upper limb motor recovery.</td>
</tr>
<tr>
<td>7</td>
<td>Laver et al [16], 2015</td>
<td>Virtual reality for stroke rehabilitation</td>
<td>27</td>
<td>0.02</td>
<td>This paper assessed the effectiveness of VR rehabilitation training.</td>
</tr>
<tr>
<td>8</td>
<td>Joo et al [20], 2010</td>
<td>A feasibility study using interactive commercial off-the-shelf computer gaming in upper limb rehabilitation in patients after stroke</td>
<td>26</td>
<td>0.02</td>
<td>The aim of this study was to assess the feasibility of using the Nintendo Wii as an adjunct to conventional rehabilitation for patients.</td>
</tr>
<tr>
<td>9</td>
<td>Adamovich et al [21], 2009</td>
<td>Sensorimotor training in virtual reality: A review</td>
<td>25</td>
<td>0.01</td>
<td>This paper discussed possible underlying mechanisms in the area of VR rehabilitation and provided some development direction.</td>
</tr>
<tr>
<td>10</td>
<td>Holden [22], 2005</td>
<td>Virtual Environments for Motor Rehabilitation: Review</td>
<td>25</td>
<td>0.05</td>
<td>In this paper, the state of VR applications in the field of motor rehabilitation was reviewed.</td>
</tr>
</tbody>
</table>
Analysis of Co-occurrence and Burst Keywords

Keywords in similar publications were identified and processed. The top 20 keywords, which had a high link strength, are outlined in Table 5. As well as identifying thematic areas in this research field, our keyword examination of all articles (N=1617) identified 50 keywords with at least 16 occurrences (Figure 8). We investigated hot spot shifts from a temporal perspective, using the top 2 keywords with the strongest citation burst—video games (2017-2021) and young adults (2018-2021; Figure 9).
Table 5. Total link strength of the top 20 occurring keywords.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keywords</th>
<th>Occurrences, n</th>
<th>Total link strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>virtual reality</td>
<td>543</td>
<td>1305</td>
</tr>
<tr>
<td>2</td>
<td>rehabilitation</td>
<td>318</td>
<td>924</td>
</tr>
<tr>
<td>3</td>
<td>stroke</td>
<td>242</td>
<td>671</td>
</tr>
<tr>
<td>4</td>
<td>vocational rehabilitation</td>
<td>85</td>
<td>115</td>
</tr>
<tr>
<td>5</td>
<td>balance</td>
<td>82</td>
<td>232</td>
</tr>
<tr>
<td>6</td>
<td>cerebral palsy</td>
<td>78</td>
<td>205</td>
</tr>
<tr>
<td>7</td>
<td>gait</td>
<td>54</td>
<td>160</td>
</tr>
<tr>
<td>8</td>
<td>upper limb</td>
<td>42</td>
<td>138</td>
</tr>
<tr>
<td>9</td>
<td>upper extremity</td>
<td>41</td>
<td>143</td>
</tr>
<tr>
<td>10</td>
<td>video games</td>
<td>41</td>
<td>138</td>
</tr>
<tr>
<td>11</td>
<td>employment</td>
<td>37</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>technology</td>
<td>37</td>
<td>75</td>
</tr>
<tr>
<td>13</td>
<td>augmented reality</td>
<td>37</td>
<td>69</td>
</tr>
<tr>
<td>14</td>
<td>traumatic brain injury</td>
<td>35</td>
<td>94</td>
</tr>
<tr>
<td>15</td>
<td>exercise</td>
<td>32</td>
<td>98</td>
</tr>
<tr>
<td>16</td>
<td>postural balance</td>
<td>31</td>
<td>96</td>
</tr>
<tr>
<td>17</td>
<td>telerehabilitation</td>
<td>29</td>
<td>84</td>
</tr>
<tr>
<td>18</td>
<td>cognition</td>
<td>29</td>
<td>82</td>
</tr>
<tr>
<td>19</td>
<td>disability</td>
<td>28</td>
<td>52</td>
</tr>
<tr>
<td>20</td>
<td>motor learning</td>
<td>27</td>
<td>96</td>
</tr>
</tbody>
</table>

Figure 8. The network map of keywords. This shows 50 keywords and is divided into 9 clusters.
Figure 9. The top 25 keywords with the strongest citation bursts from publications on virtual reality rehabilitation.

### Discussion

#### General Data

This study analyzed 1617 SCIE papers that were published between 2000 and 2021 and were related to the research of VR rehabilitation. The United States published the highest number of studies (663/1617, 41.01%), and Canada published the second highest number of studies (143/1617, 8.84%). Most of the core research institutions in this field are from the United States and Israel. The US Food and Drug Administration's encouragement of digital therapy innovation is the main reason that the largest number of publications is in North American countries, led by the United States. Israel encourages entrepreneurship in the high-tech industry; as such, 4 of the top 10 universities in this field are in Israel. The most widely read journal was *Journal of NeuroEngineering and Rehabilitation*. These findings showed that *Journal of NeuroEngineering and Rehabilitation* significantly contributed to research in this area. *Journal of NeuroEngineering and Rehabilitation* focuses on the publication of research results in the fields of neuroscience, biomedical engineering, and rehabilitation, which is very much in line with the publication of papers in this field and has attracted the attention of relevant researchers. We also examined the top 10 cited publications; the top article by Saposnik et al [14] appeared in *Stroke* and was cited 58 times.

#### The Knowledge Base and Current Research Characteristics

In previous studies, different VR-assisted rehabilitation systems were investigated in this field, with remarkable results. As

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Year</th>
<th>Strength</th>
<th>Begin</th>
<th>End</th>
<th>2000 - 2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>african american</td>
<td>2000</td>
<td>6.74</td>
<td>2000</td>
<td>2010</td>
<td>✅</td>
</tr>
<tr>
<td>service</td>
<td>2000</td>
<td>5.35</td>
<td>2000</td>
<td>2010</td>
<td>✅</td>
</tr>
<tr>
<td>environment</td>
<td>2000</td>
<td>10.1</td>
<td>2001</td>
<td>2011</td>
<td>✅</td>
</tr>
<tr>
<td>cognitive rehabilitation</td>
<td>2000</td>
<td>5.28</td>
<td>2002</td>
<td>2013</td>
<td>✅</td>
</tr>
<tr>
<td>brain injury</td>
<td>2000</td>
<td>5.03</td>
<td>2003</td>
<td>2011</td>
<td>✅</td>
</tr>
<tr>
<td>reality</td>
<td>2000</td>
<td>4.08</td>
<td>2005</td>
<td>2010</td>
<td>✅</td>
</tr>
<tr>
<td>arm</td>
<td>2000</td>
<td>9.73</td>
<td>2006</td>
<td>2012</td>
<td>✅</td>
</tr>
<tr>
<td>work</td>
<td>2000</td>
<td>5.55</td>
<td>2007</td>
<td>2012</td>
<td>✅</td>
</tr>
<tr>
<td>pattern</td>
<td>2000</td>
<td>4.1</td>
<td>2007</td>
<td>2012</td>
<td>✅</td>
</tr>
<tr>
<td>feedback</td>
<td>2000</td>
<td>4.08</td>
<td>2007</td>
<td>2014</td>
<td>✅</td>
</tr>
<tr>
<td>cerebral palsy</td>
<td>2000</td>
<td>4.11</td>
<td>2009</td>
<td>2012</td>
<td>✅</td>
</tr>
<tr>
<td>activation</td>
<td>2000</td>
<td>3.85</td>
<td>2009</td>
<td>2013</td>
<td>✅</td>
</tr>
<tr>
<td>return</td>
<td>2000</td>
<td>5.04</td>
<td>2010</td>
<td>2014</td>
<td>✅</td>
</tr>
<tr>
<td>low cost</td>
<td>2000</td>
<td>4.37</td>
<td>2011</td>
<td>2015</td>
<td>✅</td>
</tr>
<tr>
<td>low back pain</td>
<td>2000</td>
<td>3.71</td>
<td>2011</td>
<td>2015</td>
<td>✅</td>
</tr>
<tr>
<td>wii</td>
<td>2000</td>
<td>5.74</td>
<td>2012</td>
<td>2014</td>
<td>✅</td>
</tr>
<tr>
<td>randomized clinical trial</td>
<td>2000</td>
<td>4.95</td>
<td>2014</td>
<td>2015</td>
<td>✅</td>
</tr>
<tr>
<td>wii fit</td>
<td>2000</td>
<td>4.92</td>
<td>2014</td>
<td>2017</td>
<td>✅</td>
</tr>
<tr>
<td>randomized controlled trial</td>
<td>2000</td>
<td>6.12</td>
<td>2015</td>
<td>2018</td>
<td>✅</td>
</tr>
<tr>
<td>older adult</td>
<td>2000</td>
<td>4.73</td>
<td>2015</td>
<td>2018</td>
<td>✅</td>
</tr>
<tr>
<td>stroke patient</td>
<td>2000</td>
<td>3.93</td>
<td>2016</td>
<td>2018</td>
<td>✅</td>
</tr>
<tr>
<td>video game</td>
<td>2000</td>
<td>4.74</td>
<td>2017</td>
<td>2021</td>
<td>✅</td>
</tr>
<tr>
<td>young adult</td>
<td>2000</td>
<td>4.5</td>
<td>2018</td>
<td>2021</td>
<td>✅</td>
</tr>
</tbody>
</table>
indicated in Figure 7, after cocited reference clustering, key clustering nodes identified knowledge bases in this area, as follows: cluster 0 (kinematics), cluster 1 (neurorehabilitation), cluster 2 (brain injury), cluster 3 (exergame), cluster 4 (aging), cluster 5 (motor rehabilitation), cluster 6 (mobility), cluster 7 (cerebral palsy), and cluster 8 (exercise intensity). To this end, we described VR rehabilitation research knowledge bases by using different clusters, with time considerations.

For cluster 0, the kinematics cluster, scientists designed a VR rehabilitation system based on kinematics theory to measure, monitor, and predict outcomes; they achieved good results. For example, researchers developed various VR scenes for rehabilitation training for cervical vertebrae injuries; lower limb injuries of athletes; and arm, hand, and trunk movements in patients with stroke. They proposed kinematic evaluation and measurement methods based on the VR environment [24-26].

In cluster 1, the neurorehabilitation cluster, the research interest in VR technology for neurorehabilitation is increasing. Scientists have experimentally demonstrated the role of VR systems in neural rehabilitation and motor assistance from the perspectives of patients with impaired visuospatial perception after stroke, the augmented effect of intermittent theta burst stimulation on neurorehabilitation programs, and the use of motor imagination in brain-computer interfaces. Therefore, VR programs are considered safe and can be performed with standard neurorehabilitation protocols in patients with neurological conditions [27-29].

In cluster 2, the brain injury cluster, scientists demonstrated the experimental results of VR training systems for patients with brain injury, for those with acute onset central nervous system damage, and for vocational rehabilitation training from both the experimental perspective and the retrospective perspective. Importantly, most patients were responsive to VR training and showed improvements in gait function, balance control, vocational rehabilitation training, and other aspects [30-32].

In cluster 3, the exergame cluster, McMahon et al [33] investigated VR exergaming to augment physical activity in students (high school) with intellectual and developmental disabilities. Their findings indicated that when students used the VR exercise exergaming intervention, they increased the duration and intensity of their physical activity. Nambi et al [34], in their psychological and hormone analysis study, discovered that VR exergaming was effective for American soccer players with chronic low back pain when compared with conventional exercise training programs. In a multicenter controlled trial, Meyns et al [35] demonstrated that exergaming in children with spastic cerebral palsy improved balance after previous poor balance performances.

In cluster 4, the aging cluster, several investigations concentrated on upper limb rehabilitation intervention and overall skill improvement for older patients. Molina et al [36] analyzed the impacts of 13 VR sports games on physical function in older patients. Another group investigated the elbow and shoulder movements of older patients with chronic stroke, using the Predict Recovery Potential algorithm. The training improved motivation, but the benefits of physical function in aging were unknown [2,37].

In cluster 5, the motor rehabilitation cluster, scientists investigated various VR training devices. A team, for example, used VR-enhanced robot-assisted gait training to monitor the gait, motion, balance, fear of falling, and independence of 15 patients with chronic stroke. The researchers observed that cognitive performance and gait speed in tasks improved among study participants (P<.05) [38]. Maier et al [39], in their systematic investigation on randomized controlled trials, identified the benefits of specific VR systems for rehabilitating upper limb function and activity after stroke and reported that specific VR systems were highly beneficial for upper limb recovery when compared with conventional therapies.

In cluster 6, the mobility cluster, researchers designed remote VR and augmented reality rehabilitation training treadmills to record training data from patients with Parkinson disease or stroke. The experimental results showed that flexibility and balance improved and that VR rehabilitation methods could provide personalized rehabilitation strategies while also improving patient participation [40,41]. To identify the benefits of a VR-based interventional therapy, a previous scoping review reported that dynamic balance measures improved significantly following a therapeutic intervention; therefore, robust study designs that focus on intensity and dose responses from VR training could improve the efficacy of methods for treating mobility disorders [42].

In cluster 7, the cerebral palsy cluster, scientists studied gait, spatial perception exercises, functional mobility, exoskeletons, visuomotor construction, and other aspects of cerebral palsy in children by using data obtained through a VR rehabilitation training system. Bimanual performance and cognitive rehabilitation gains improved significantly during gait rehabilitation [43-45]. Furthermore, an augmented reality real-time feedback approach that incorporated infrared recognition technology was used to assess temporal and gait parameters during gait training in children with cerebral palsy. Importantly, cadence, velocity, functional ambulation, and bilateral step and stride length were augmented after the intervention [46].

For cluster 8, the exercise intensity cluster, scientists discovered that training factors (training intensity, prolonged activity, and rest were optimized) could potentially provide training optimization stimuli. de Vries et al [47] identified some elements of VR balance games that could potentially provide strength training stimulation. Baniña et al [48] investigated the relationship between upper limb motor recovery and exercise intensity in participants with subacute stroke. Importantly, they found that VR rehabilitation could be used to generate intensive exercise programs; however, clearer exercise progression guidelines should be published.

Scientific Frontiers and Future Research Trajectories

Keywords represent current research issues or topical concepts, while emerging trends and research frontiers are represented by burst keywords. We used the CiteSpace program to gather burst keywords and identified 2 scientific frontier areas with the strongest citation bursts—video games (2017-2021) and young adults (2018-2021).
**Video Games**

We found that VR video games are simple and appealing to patients and therefore encourage patient participation. Measurements and assessments that were taken after testing VR video games with 12 older patients and 13 children with intellectual disabilities showed clinical improvements in motor capabilities and a positive effect on quality of life [49,50]. Jung et al [51] investigated the impact of Xbox Kinect (Microsoft Corporation) training on lower extremity motor function in adolescents with spastic diplegia cerebral palsy. The outcome measurements, which were measured by using standard industry tools, showed that adolescents displayed significant improvements. A previous randomized controlled trial assessed the impact of a VR dance training program, which involved using an Xbox 360 Kinect game, on kyphosis angle and respiratory parameters in young women with postural hyperkyphosis and showed that dance games were an effective therapy [52]. Deutsch et al [53] asked 15 individuals with mild to moderate lower extremity deficits in the chronic poststroke phase to use the Kinect Light Race game and showed that video games provided comparable intensity, improved accuracy, greater enjoyment, and less exertion when compared with standard care activities. The application of VR-based video games in the field of rehabilitation is the current potential research direction.

**Young Adults**

VR rehabilitation training research has primarily focused on sport rehabilitation aspects, such as the spine, gait, or balance, with fewer studies on teenagers when compared to studies on adults. The burst keyword *young adult* appeared between 2018 and 2021, indicating that scholars have begun to pay attention to the positive physiological, social, and psychological impacts on teenagers. Fralish et al [54], in a case study of 6 VR sessions over a 3-week period, reported that specific VR programs were putatively physiologically and psychosocially helpful (mood improvement) for young adults with physical disabilities. Smith et al [55] proposed a VR job interview training system for participants with severe mental illness and transition-aged youth with autism spectrum disorders, which resulted in improved interview skills and employment access.

**Conclusions**

In this study, we used a bibliometric analysis to objectively, comprehensively, and systematically analyze the VR rehabilitation research literature. We identified the knowledge bases, current topical hot spots, and oncoming trends in this area. In this research field, the areas with acceptable knowledge bases included kinematics, neurorehabilitation, brain injury, exergaming, aging, motor rehabilitation, mobility, cerebral palsy, and exercise intensity. We theorize that future emerging trends and frontiers will focus on video games and young adults. We also identified contemporary research topics and future trends in VR rehabilitation research and provided guidance for future research in this exciting area.

Although papers that were published at different times were gathered for this study, some were not comprehensive in nature and may have introduced publication bias, thereby affecting study outcomes.

**Limitations**

Our study still has some limitations to be addressed. Although the WSCC database is reliable, other databases, such as non-English databases, should also be considered to ensure that all relevant papers are collected. Additionally, the rapid progress in the field of VR technology application limits the timeliness of this bibliometric study. Further, the VR-based rehabilitation activities we investigated included some gaming exercises, exergames, and video games that were not clearly defined. Finally, the results of the included studies only reflect current research trends in academia. In the future, VR technology is expected to be further integrated with clinical applications, which will enrich and improve the research in this subject field.

**Acknowledgments**

This study was supported by grants from the Association of Fundamental Computing Education in Chinese Universities, which were awarded to ZZ (grant 2019-AFCEC-199). Further, TF was awarded a grant for a research project titled “Research on the application of genetic big data analysis in the prevention and control of newborn birth defects” (grant 2900021013-CMU-013).

**Conflicts of Interest**

None declared.

**References**


**Abbreviations**

SCIE: Science Citation Index Expanded
VR: virtual reality
WSCC: Web of Science Core Collection

©Ting Fan, Xiaobei Wang, Xiaoxi Song, Gang Zhao, Zhichang Zhang. Originally published in JMIR Serious Games (https://games.jmir.org), 06.03.2023. 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.
The Effects of Exergaming on Attention in Children With Attention Deficit/Hyperactivity Disorder: Randomized Controlled Trial

HongQing Ji1,2, PhD; Shanshan Wu1,2, PhD; Junyeon Won3, PhD; Shiyang Weng1, MSPE; Sujin Lee2, MEd; Sangmin Seo4, MD; Jung-Jun Park2, PhD

1School of Physical Education & Health, Wenzhou University, Wenzhou, China
2Division of Sport Science, Pusan National University, Busan, Republic of Korea
3Institute for Exercise and Environmental Medicine, Texas Health Presbyterian Hospital, Dallas, TX, United States
4Busan Children's Mind Clinic, Busan, Republic of Korea

Corresponding Author:
Jung-Jun Park, PhD
Division of Sport Science
Pusan National University
63 beon-gil 2 Busandaehak-ro
Geumjeong-gu
Busan, 46241
Republic of Korea
Phone: 82 051 510 2713
Fax: 82 051 510 3746
Email: jjparkpnu@pusan.ac.kr

Abstract

Background: Despite growing evidence showing the effects of exercise and cognitive trainings on enhancing attention, little is known about the combined effects of exergame on attention in children with attention deficit/hyperactivity disorder (ADHD). Exergame, a form of exercise using a video game, has both cognitive stimulation and physical activity components and has been shown to improve cognitive function in children.

Objective: The purpose of this study was to investigate the effect of exergaming on attention and to compare the effect induced by exergaming with the effect of aerobic exercise on attention in children with ADHD.

Methods: In all, 30 children with ADHD, aged 8-12 years, were randomly divided into an exergaming group (EXG; n=16) or a bicycle exercise group (BEG; n=14). Before and after the 4-week intervention, the Frankfurter Aufmerksamkeits-Inventar (FAIR; Frankfurt Attention Inventory) test was administrated, and event-related potentials during the Go/No-go task was measured to assess attention.

Results: After intervention, both the EXG and BEG had significantly increased selective attention and continuous attention (all \(P<.001\)), as well as self-control on the FAIR test (EXG: \(P=.02\) and BEG: \(P=.005\)). Similarly, both the EXG and BEG had significantly reduced response time on the Go/No-go test (all \(P<.001\)). For the Go response, the N2 amplitude (frontocentral maximal negativity) was significantly increased in Fz (midfrontal line) in the EXG (\(P=.003\)) but was not changed in the BEG (\(P=.97\)). Importantly, the N2 amplitude in Fz was significantly greater in the EXG compared to the BEG (Go: \(P=.001\) and No-go: \(P=.008\)).

Conclusions: Exergaming has the comparable effects to bicycle exercise to enhance attention in children with ADHD, suggesting that exergaming can be used as an alternative treatment for children with ADHD.

Trial Registration: Clinical Research Information Service KCT0008239; https://tinyurl.com/57e4jtnb

(JMIR Serious Games 2023;11:e40438) doi:10.2196/40438

KEYWORDS
exergame; N2 amplitude; attention function; response time; attention deficit/hyperactivity disorder; ADHD
Introduction

Attention deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is characterized by inattention, hyperactivity, and impulsivity, or both [1]. Approximately 3% to 5% of the school-age population are living with ADHD [2,3]. Among symptoms of ADHD, hyperactivity tends to improve during adolescence, but carelessness and impulsiveness may persist into adulthood [4]. The impaired attention and executive function in children with ADHD lead to reduced inhibitory ability, working memory, and task shift due to an imbalance of neurotransmitters such as dopamine and norepinephrine [5-7]. The abnormal neuropsychological activities caused by these symptoms may disrupt the patients’ learning ability, daily activities, professional activities, and social functions [8,9].

In children with ADHD (aged >8 years), dysfunctions in neuronal networks associated with attentional processes and cognitive control have been well documented using event-related potentials (ERPs) [10-13]. ERP studies used tasks (eg, Eriksen flanker and Go/No-go tasks) to assess both behavioral and neurophysiological perspectives of ADHD [14]. For example, behavioral performance (eg, response time [RT]) is commonly used to determine processing speed [15], and neurophysiological technology has been used to measure selective attention capacity and skills [16]. The No-go N2 components have been reported to be abnormally lower in children with ADHD than those in children with other neurodevelopmental disorders [17,18]. The N2 amplitude occurs for about 170-350 milliseconds and is related to attention and cognitive control, specifically an inhibitory control response [16,19,20]. Particularly, the amplitude of the N2 is an indication of the intensity of information processing required for the discrimination of stimuli, and the latent phase of the N2 reflects the time of the cognitive processing that distinguishes sensory stimuli [21].

Exercise has been known to enhance attention, executive function, impulsivity, and hyperactivity in children with ADHD by promoting the secretion of neurotransmitters (eg, dopamine and norepinephrine) in the brain, thereby inducing positive responses on the main symptoms of ADHD [22,23]. Although most of the evidence has focused on the acute effects of aerobic exercise on RT and P3 components of children with ADHD, N2 components have received comparatively little attention. Furthermore, there are mixed results in the study of exercise and N2 in normal young adults and preadolescent children [24-26]. For example, Ligeza et al [24] reported that acute moderate intensity exercise increased the conflict effect of N2 amplitude in young people, which suggested improvement in inhibition after a single session of moderate exercise compared to high-intensity interval exercise and seated rest condition. In another study, acute moderate-intensity exercise led to reduced N2 amplitude in preadolescent children and young adults, suggesting that acute exercise might have a general effect on the conflict detection process [25,26]. Therefore, further studies are needed to explore whether exercise can improve the N2 components of children with ADHD.

However, previous studies reported that cognitively engaging physical exercise leads to benefits for cognitive performance [27,28]. For example, a combination of physical activity and cognitive stimulation improved reaction times in inhibition and switching neural efficiency [27,28]. Exergaming is a video game that requires body movement and provides active gaming experience as a form of physical activity [29]. Due to motivational issues and a low level of positive reinforcement, traditional (cognitive) training programs are often uninteresting and fatiguing for children with ADHD [30]. Exergaming helps combine physical and cognitive training through an intriguing game. Recently, exergaming has attracted great attention because of their growing popularity among children, adolescents, and older adults [31-35] due to its effectiveness in motivating participants to engage in intervention [36].

Recent studies revealed that exergaming improves the executive function in healthy adolescents by increasing cognitive stimulation [31]. Exergaming has also been reported to enhance executive function in children with ADHD compared to the control group [37,38]. However, it is currently in question whether these effects are caused by games or exercise. Moreover, there is, to date, only one study that has measured the effects of exergaming on N2 amplitude. This study showed that exergaming led to greater aerobic capacity and larger N2 amplitude, an indication of improved selective attention in patients with metabolic syndrome [39]. Despite the evidence regarding the effects of exergaming on cognitive function, how the neural processing of N2 components is altered in children with ADHD after an exergaming intervention remains unclear. Moreover, it remains inconclusive if exergaming improves attention in children with ADHD and whether exergaming has additional benefits compared to traditional form of aerobic exercise. Therefore, the purpose of this study was to investigate the effect of exergaming on attention and to compare the effect induced by exergaming with the effect of aerobic exercise on attention in children with ADHD. Our first hypothesis was that exergaming will result in improved Go/No-go task performance and increased N2 amplitude in children with ADHD. Second, we hypothesized that the improvement in Go/No-go task performance and N2 amplitude will be greater in exergaming compared to bicycle exercise.

Methods

Study Design and Participants

We used a randomized controlled trial in which participants were randomly assigned to either the exergaming group (EXG) or the bicycle exercise group (BEG). Children with mild to moderate ADHD were recruited from the Children’s Mental Health Medical Center in Busan, Republic of Korea. Participants who met the following criteria were included in the study: a Korean Attention-Deficiit/Hyperactivity Disorder Diagnostic Scale score between 70-110, indicating a lack of attention; the absence of diseases other than ADHD; right-hand dominance; normal or corrected-to-normal vision; and the absence of physical impairment to perform exercise. In all, 42 children (aged 8-12 years) with mild to moderate ADHD were recruited and then randomly assigned to either the EXG (n=21) or BEG (n=21). A total of 12 participants (5 for the EXG and 7 for the BEG) were withdrawn from the study.
resulting in the final number of 30 participants in this study (16 in the EXG and 14 in the BEG; Figure 1). All participants maintained the same medication and dose during the exercise intervention.

Figure 1. Flowchart of participant eligibility, withdrawals, and the final sample included in the final analysis. BEG: bicycle exercise group; EXG: exergaming group.

**Ethics Approval**
Written informed consent was obtained from both the children and their parents or legal guardians. This study was approved by the Pusan National University Institutional Review Board in accordance with the Helsinki Declaration (PNU IRB/2018_30_HR).

**Sample Size**
The sample size was calculated using a sample size calculation software program (G*Power, version 3.1.9.2 for Windows), with an effect size of 0.58, statistical power of .80, and statistical level of significance of .05. The effect size was calculated from previous studies [40]. As a result, the sample size for each group was established at 15 patients, so we decided to recruit 42 patients for each group in consideration of a potential 30% dropout rate.

**Exergaming and Bicycle Exercise Intervention**
The EXG was administered using ExerHeart devices (D&J Humancare), which consisted of a running or jumping board (730-cm width × 730-cm depth × 130-cm height) and a screen connected to the board (Figure 2A). The exercise program called *Alchemist’s Treasure* (D&J Humancare) was used for the EXG. In this game, participants run or jump in place with their avatars, using the front, back, left, and right sensors on the mat to avoid obstacles and acquire items (Figure 2B; Multimedia Appendix 1). The BEG performed stationary bike exercise using commercial Fit Elite-Whole body exerciser 1000, with resistance of 0.5~3 kiloponds.

Exercise session for both the EXG and BEG consisted of 3 days/week, 50 min/day, and 60% to 80% of heart rate (HR) reserve for 4 weeks. We monitored individual exercise intensity using an HR monitor (Polar RS400sd). Exercise intensity was determined by using the Karvonen target HR method: exercise intensity × (HR_max – resting HR) + resting HR [41]. Each exercise session consisted of 10 minutes of warm-up, 30 minutes of main exercise, and 10 minutes of cooldown. Exercise interventions were conducted at 2 separate child mental health care facilities so that the groups were not aware of the exercise program they were performing. Participants who did not complete more than 80% of the exercise sessions were excluded from the final analysis.

For both the EXG and BEG, all subjects’ HRs during exercise were monitored using HR monitors (Polar RS400sd) to confirm that the values were within the target HR range.
Go/No-go Task

The Go/No-go task was used to measure the capacity to sustain attention and response control. We used the computerized Go/No-go task using a built-in software of the electroencephalography (EEG) analyzer (Telescan; LAXTHA). During the Go/No-go task, each participant was instructed to press a button for a frequent stimulus (ie, Go stimulus, 60% probability) or to withhold a response for an infrequent stimulus (ie, No-go stimulus, 40% probability). The stimulus was presented in the center of a 15-inch screen with a white background, and the size was 3 × 3 cm. In each trial, a signal consisting of an interstimulus interval was first presented for 1000 milliseconds to attract the participant’s attention. This signal was followed by either the Go stimulus (ie, lion image), No-go stimulus (ie, tiger image), or Control stimulus (ie, leopard image) for 1500 milliseconds. Participants were instructed to press the left arrow keyboard for the Go stimulus and the right arrow keyboard for the No-go stimulus and not to respond to the Control stimulus. Participants underwent a familiarization session before starting the test. The number of practice trials was standardized (1 trial each) so that participants performed the same number of practice trials before each condition. The response time (milliseconds) and the accuracy rate (% correct) were recorded simultaneously from the start of the reaction. The entire task (including practice trials) lasted for 10 minutes.

EEG Measurements

EEG was measured from 31 Ag/Ag–CL electrodes, placed according to an augmented 10–20 International System. All EEG recordings were referenced to the average of the right and left mastoid, and the ground electrode was placed on the Fz electrode site. The electrooculogram activity was recorded from electrodes attached below and above the left eye and electrodes located at the outer canthi of both eyes. All electrodes were maintained at impedances <10 kΩ before data recording. After the completion of data collection, EEG signals were analyzed using software (Telescan; LAXTHA); the bandpass filter of the amplifier was 0.5-20 Hz, the sampling rate was 1000 Hz, and a notch filter was at 60 Hz.

The stimulus-locked epochs acquired for the Go/No-go test were extracted offline from 200 milliseconds before to 1500 milliseconds after the stimulus onset, and the period from –100 to 0 milliseconds before stimuli onset was used as the baseline. Peak amplitudes and latencies were measured automatically. The N2 peak amplitudes and latencies were measured in the difference waves of the attended condition. The N2 component was defined as the largest positive peak occurring between 170-320 milliseconds after stimuli. EEG data were collected before and after 8 weeks of the exercise intervention [42,43].

Frankfurter Aufmerksamkeits-Inventar

The Frankfurter Aufmerksamkeits-Inventar (FAIR; Frankfurt Attention Inventory) is a psychological test to assess attention and concentration [44]. The FAIR tests the ability to quickly
distinguish correct signal among many other similar signals and to hide insignificant information (ie, external interference). Two different versions of the FAIR test were used for test reliability. A total of 640 test items were divided into tests 1 and 2. Each FAIR test paper was arranged in the shape of 20 horizontal and 16 vertical lines on a single sheet of paper, and there were a total of 320 test items. Participants were instructed to find 2 correct shapes, such as a “three-point circle” and “two-point square” among the 4 shapes and draw a line from left to right with a pencil on the test paper to mark them. The scoring was administered using the perspective of 3 attentional actions, and details are shown in Table 1. The ability index P is an index of the number of items related to attention during the test. The control index Q represents the proportional value of correct judgment as a result of attention among all responses. The persistence index C indicates how consistently the attention task was maintained.

Table 1. Frankfurter Aufmerksamkeits-Inventar inspection dimensions.

<table>
<thead>
<tr>
<th>Test item</th>
<th>Related ability</th>
<th>Scoring method</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Selective attention</td>
<td>((T^b - EL^c) - 2(EO^d + EC^e))</td>
<td>.944</td>
</tr>
<tr>
<td>Q</td>
<td>Self-control</td>
<td>(P ÷ T)</td>
<td>.903</td>
</tr>
<tr>
<td>C</td>
<td>Sustained attention</td>
<td>(P \times Q)</td>
<td>.941</td>
</tr>
</tbody>
</table>

aP: performance value.  
bT: total number of items worked.  
cEL: total number of line drawing errors.  
dEO: total number of teeth not marked on a target item.  
eEC: total number of teeth marked that are not a target item.  
fQ: quality value.  
gC: continuity value.

Statistical Analysis
The normality of Go/No-Go test performance was also tested using the Shapiro-Wilk test. Repeated measures ANOVA were used to compute the main effects of time (ie, before vs after intervention), group (ie, EXG vs BEG), and group \times time interaction on the behavioral performance (ie, accuracy rate and reaction time), ERP (ie, N2 amplitude), and FAIR (ie, ability index P, control index Q, and persistence index C). If an interaction was identified, paired sample \(t\) test (2-tailed) was used to verify the direction of the interaction. Significance level was set at .05 for all analyses, effect sizes were assessed using partial eta squared (\(\eta^2_p\)), and all statistical analyses were performed using SPSS (version 24; IBM Corp).

Results

Participants
Of the 42 children who participated in this study, 5 in the EXG dropped out of the study due to conflicting schedule and 7 in the BEG dropped out of the study due to lost interest. Demographic and physical characteristics for all subjects are provided in Table 2.

Table 2. Summary of the participant demographic information.

<table>
<thead>
<tr>
<th></th>
<th>EXG^a (n=16)</th>
<th>BEG^b (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>9.00 (1.46)</td>
<td>8.85 (1.63)</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>14 (88)</td>
<td>12 (86)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>129.38 (9.68)</td>
<td>134.69 (7.76)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>31.66 (11.84)</td>
<td>35.31 (6.84)</td>
</tr>
<tr>
<td>BMI (kg/m^2), mean (SD)</td>
<td>18.5 (5.4)</td>
<td>19.3 (2.7)</td>
</tr>
<tr>
<td>K-ADHDDS^c (index), mean (SD)</td>
<td>96.14 (19.16)</td>
<td>87.50 (13.17)</td>
</tr>
</tbody>
</table>

aEXG: exergaming group.  
bBEG: bicycle exercise group.  
cK-ADHDDS: Korean Attention-Deficit/Hyperactivity Disorder Diagnostic Scale.

Behavioral Indices
There was no significant group \times time interaction on the changes in RT in response to Go and No-go stimulations before and after exercise intervention (\(P=.65\) and \(P=.807\), respectively). However, both groups significantly reduced RT for Go and No-go stimulations after exercise intervention (Figure 3). For Go stimulation, RT was significantly reduced in the EXG from 846.13 (SD 126.43) milliseconds to 764.56 (SD 107.05) milliseconds (\(P=.001; \eta^2_p=0.969\)) and in the BEG from 873.79 (SD 123.31) milliseconds to 780.79 (SD 96.41) milliseconds.
(\(P=.001; \eta^2_p=1.267\); Figure 3A). For No-go stimulation, RT was significantly reduced in the EXG from 872.25 (SD 96.53) milliseconds to 786.75 (SD 100.02) milliseconds (\(P=.001; \eta^2_p=1.211\)) and in the BEG from 903.71 (SD 130.78) milliseconds to 813.77 (SD 111.86) milliseconds (\(P=.001; \eta^2_p=1.114\); Figure 3B).

**Figure 3.** Reaction time during Go/No-go task in the EXG and BEG before and after exercise intervention (A) Go stimulus, (B) No-go stimulus. BEG: bicycle exercise group; EXG: exergaming group. **P**<.001.

### N2 Amplitude

There was a significant group × time interaction on the Go and No-go N2 amplitudes in the Fz region before and after exercise intervention (\(P=.049\) and \(P=.047\), respectively; Table 3). The EXG had significantly increased Go N2 amplitude in Fz after intervention (\(P=.003\)), whereas the BEG showed no significant changes (\(P=.97\)). Neither group showed changes in No-go stimulus after intervention. In the between-group comparison, the EXG consistently demonstrated greater N2 amplitudes on both Go and No-go stimulations. The waveforms of Go and No-go N2 amplitudes in Fz for the EXG and BEG before and after exercise are shown in Figure 4.

### Table 3. Results of changes in Go/No-go N2 amplitude in Fz.

<table>
<thead>
<tr>
<th>Effects and group</th>
<th>Preintervention</th>
<th>Postintervention</th>
<th>(t) test</th>
<th>(P) value</th>
<th>(\eta^2_p)</th>
<th>G(^a)</th>
<th>T(^b)</th>
<th>G×T(^c)</th>
<th>(\eta^2_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Go stimulus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXG(^d), mean (SD)</td>
<td>−3.92 (3.78)</td>
<td>−6.82 (4.49)</td>
<td>.003</td>
<td>1.791</td>
<td>.02</td>
<td>.054</td>
<td>.049</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>BEG(^e), mean (SD)</td>
<td>−2.33 (2.15)</td>
<td>−2.30 (4.05)</td>
<td>.97</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>t test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P) value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta^2_p)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.503</td>
<td>1.045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No-go stimulus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.008</td>
<td>.68</td>
<td>.047</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>EXG, mean (SD)</td>
<td>−4.64 (3.65)</td>
<td>−6.70 (3.85)</td>
<td>.09</td>
<td>0.911</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEG, mean (SD)</td>
<td>−3.55 (3.91)</td>
<td>−2.18 (2.36)</td>
<td>.27</td>
<td>0.615</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>t test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.44</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(P) value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.289</td>
<td>1.345</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta^2_p)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)G: group effect.  
\(^b\)T: time effect.  
\(^c\)GoT: interaction effect between group and time.  
\(^d\)EXG: exergaming group.  
\(^e\)BEG: bicycle exercise group.
Figure 4. Average event-related potential waveforms of Fz for mean N2 amplitudes during the Go/No-go test: (A) Go stimulus and (B) No-go stimulus in the exergame group (EXG) and bicycle exercise group (BEG) before and after the exercise intervention.

Aufmerksamkeits-Inventar Data

There was no significant group × time interaction on the ability index P (selective attention) before and after exercise intervention ($P=.79$). However, both groups significantly increased ability index P after exercise intervention (both $P<.001$). Similarly, although there was no significant interaction between the EXG and BEG on the changes in ability index Q (self-control; $P=.68$), both the EXG and BEG significantly increased control index Q following intervention (EXG: $P=.02$ and BEG: $P=.005$). Lastly, there was also no significant interaction between the EXG and BEG on the changes in ability index C (persistent attention; $P=.66$), but both groups significantly increased index C in response to exercise intervention (both $P<.001$; Table 4).
### Table 4. Results of changes in Frankfurter Aufmerksamkeits-Inventar (FAIR).

<table>
<thead>
<tr>
<th>Effects and group</th>
<th>Preintervention</th>
<th>Postintervention</th>
<th>$t$ test</th>
<th>$P$ value</th>
<th>$\eta^2_p$</th>
<th>G$^a$</th>
<th>$T^b$</th>
<th>G×T$c$</th>
<th>$\eta^2_{pG}$</th>
<th>$\eta^2_{pT}$</th>
<th>$\eta^2_{pG×T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance value (P; score)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXG$^d$, mean (SD)</td>
<td>190.44 (47.85)</td>
<td>269.06 (92.13)</td>
<td>.001</td>
<td>.19</td>
<td>.001</td>
<td>.79</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEG$^e$, mean (SD)</td>
<td>231.15 (71.44)</td>
<td>303.92 (101.78)</td>
<td>.001</td>
<td>.20</td>
<td>.001</td>
<td>.66</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$ test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td>.08</td>
<td>.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta^2_p$</td>
<td>0.670</td>
<td>0.354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality value (Q; ratio)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXG, mean (SD)</td>
<td>0.90 (0.08)</td>
<td>0.95 (0.005)</td>
<td>.02</td>
<td>.19</td>
<td>.001</td>
<td>.68</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEG, mean (SD)</td>
<td>0.92 (0.06)</td>
<td>0.97 (0.02)</td>
<td>.005</td>
<td>.20</td>
<td>.001</td>
<td>.66</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$ test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td>.31</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta^2_p$</td>
<td>0.381</td>
<td>0.451</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuity value (C; score)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXG, mean (SD)</td>
<td>171.33 (48.30)</td>
<td>262.44 (98.58)</td>
<td>.001</td>
<td>.20</td>
<td>.001</td>
<td>.66</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEG, mean (SD)</td>
<td>214.66 (76.67)</td>
<td>294.70 (103.07)</td>
<td>.001</td>
<td>.20</td>
<td>.001</td>
<td>.66</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$ test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td>.07</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta^2_p$</td>
<td>0.679</td>
<td>0.314</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$G: group effect.  
$^b$T: time effect.  
$^c$G×T: interaction effect between group and time.  
$^d$EXG: exergaming group.  
$^e$BEG: bicycle exercise group.

### Discussion

#### Principal Findings

The purpose of this study was to examine the effects of exergaming on attention in children with ADHD and whether these effects were driven by games or exercise by comparing the effects of exergaming and bicycle exercise. Through an examination of the behavioral performance and ERP during the Go/No-go task and FAIR test, our results may suggest that both exergaming and bicycle exercise elicited shorter RT during the Go/No-go task and improved selective attention, self-control, and persistent attention. Notably, larger N2 amplitudes were observed following exergaming compared to the bicycle exercise.

Our results are in agreement with prior findings that showed improved processing speed in children with ADHD after exercise. For example, acute aerobic exercise engendered significantly faster processing speed during the Stroop and flanker tasks compared to the control group [45]. In another study, an 8-week yoga program enhanced attention and processing speed in children with ADHD [46]. Research shows that exercise facilitates the upregulation of cerebral blood flow (CBF) [47] and that an exercise-induced increase in CBF promotes information processing speed [48]. Further, a possible neurophysiological mechanism underpinning the beneficial effects of exercise training on attention in children with ADHD was the exercise-induced secretion of catecholamine [49]. Therefore, in this study, it may be that both exergaming and bicycle exercise stimulate CBF increase and catecholamine secretion, thereby promoting the allocation of attention resources and improvement of processing speed in children with ADHD.

However, since this study did not measure CBF and catecholamines levels, this interpretation should be viewed with caution, and the potential neurophysiological mechanism underpinning exergaming and attention should be further investigated in the future.

Novel to this investigation is that there was an interaction effect of exercise modes on changes in the N2 amplitude during the Go/No-go task after intervention, such that the improvements in the N2 amplitude were only manifested in exergaming while the intervention effects were not observed in the bicycle exercise. The N2 components provide important information through the amplitude and latent phase because it is an indication...
of attentional assignment of the anterior cingulate cortex (ACC) when performing cognitive tasks [50,51]. Baker and Holroyd [52] suggested that task demands involving high conflict and working memory loads should strongly activate the ACC, giving rise to increased N2 amplitudes. No-go stimulation activates the prefrontal cortex and can affect cognitive regulation and inhibition processes [53]. Previous studies have reported that the N2 amplitude reflects the neural changes associated with the task [54] and that the N2 amplitude increases with greater attention [51]. Drollette et al [25] reported that the N2 amplitude decreased after running on the treadmill for 20 minutes, but another study showed that the N2 amplitude remained unchanged after an acute exercise [55]. Additionally, Stroth et al [56] pointed out that greater physical fitness is associated with better task preparation as well as decreased amplitudes in N2 amplitude, indexing more efficient executive control processes. In our previous study, both exergaming and treadmill exercise improved selective attention (N2) in patients with metabolic syndrome [39]. The results indicated that exergaming facilitated visual perceptual stimulation in the virtual environment to enhance the selective attention activity within the cerebral cortex, brain regions associated with executive function, which was not changed by normal aerobic exercise. In light of our N2 amplitude findings, exergaming is likely to better regulate the activity of ACC in the ventral prefrontal cortex than traditional form of aerobic exercise and effectively increase attention. It is hypothesized that ACC activation may have been reduced due to less-engaging exercise environment, which engendered no beneficial effect on attention after bicycle exercise. This may also be associated with the fact that all participants who dropped out from the BEG (n=7) were due to lost interest, whereas all participants who dropped out from the EXG (n=5) were due to conflicting schedule. This obvious discrepancy between the groups in the dropout reasons occurred even though the interventions were administered in 2 separate child mental health care facilities, which was intended to avoid environmental factors that could influence the participants’ motivation. Thus, the ACC is likely to be more actively stimulated during complex exercise that involves multiple aspects of environment (eg, exergaming) than monotonous exercise environment (eg, cycling). Performing exergaming requires a process of directional judgment (eg, front or back and left or right); thus, attention and cognitive control ability needs to be executed to successfully perform the task. However, further translational evidence is needed to clarify this hypothesis. Interestingly, aerobic exercise has a faster judgment of response times when it comes to information processing, but cash realization during neuroelectrical activity has a different effect. It may be that in the applied exergaming, the inhibition and switching components are needed to a greater extent than updating for successful task performance [37]. The Alchemist’s Treasure training program primarily includes strength training; coordination (and endurance); sensitivity training; as well as demands on cognitive functions such as inhibition, switching, updating, attention, and rapid action execution. Prior research has demonstrated that cognitively demanding physical activities (exergaming), such as coordination exercises, preferentially activate brain regions used to control higher-order cognitive processes, leading to improved performance [31,37]. In conclusion, exergaming has the characteristics of aerobic exercise, so it has the ability to judge reaction time quickly. Meanwhile, the brain area is stimulated by the corresponding stimulation more favorable to neuroelectrical activity than general aerobic exercise. For the FAIR results, we found improved selective attention, self-control, and persistent attention following exergaming and bicycle exercise. The results of Medina et al [57] showed that physical activity increases the release of serotonin, dopamine, and norepinephrine, and it is assumed that the attention of the participants with ADHD is improved because of the release of these neurotransmitters. According to a previous investigation, tai chi training was conducted to investigate the attentional in children and adolescents with ADHD [58]. The results showed that continuous tai chi performance at a stable speed significantly enhanced the ability of selective attention and continuous attention. Studies in adolescents with ADHD have shown that exercise-based game intervention confers maximal benefits for selective attention, self-control, and persistent attention [59,60]. We surmise that 3 components of executive function (ie, selective attention, self-control, and persistent attention) have been facilitated during exergaming performance. First, selective attention is the process of stimulating specific consciousness. During exergaming, participants had to quickly avoid the obstacles that appear on the screen; thus, participants had to maintain their selective attention throughout the game. The second component is cognitive control, which is necessary to complete tasks within a certain period of time and is expected to be facilitated while pressing the displayed number on the screen quickly and accurately during exergaming. Lastly, sustained attention is the duration of concentration related to the duration of correct concentration. In this study, exergaming characters were controlled by the user’s movement. To control the game character’s movement accurately and quickly, participants had to maintain their attention throughout the game. Therefore, because exergaming contains general aerobic exercise elements as well as additional components to stimulate selective attention, self-control, and continuous attention, it may have had significant impacts on the attention and attention of children with ADHD. Limitations Our study is not without limitations. First, this study is subject to the limitation of lacking a nonexercise (or active) control group, warranting some caution in interpreting the results until they can be replicated in a larger randomized controlled trial. Second, this study is also limited by a relatively small sample size (n=30) and homogeneous characteristics of participants (ie, all Asian and 26/30, 87% male), so the results may not be generalizable to the entire population of children with ADHD. Third, only one type of game device was used in this study; thus, future studies also need to examine the effects of multiple exergaming devices and games on ADHD attention. Fourth, the relationship between game scores and attention was not considered in this study, which should be examined in future studies.
Conclusions

Our results showed that both exergaming and bicycle exercise training enhanced attention in children with ADHD, but the benefits induced by exergaming may be somewhat greater than those from bicycle exercise. These findings suggest that, through complex and stimulating environment during the training, exergaming may stimulate the frontal lobes of the brain and has a significant impact on attention, processing speed, and cognitive control function, thereby improving attention in children with ADHD symptoms. Therefore, this study may hold an important public health implication that exergaming could be used as a new exercise therapy for children with ADHD in the future.

Acknowledgments

We acknowledge and thank Sujin Lee and Sangmin Seo for their contributions to the project. We thank the participants for their time and dedication while participating in this study.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials. 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

Participants played the video game with ExerHeart.

[MP4 File (MP4 Video), 2747 KB - games_v11i1e40438_app1.mp4 ]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1352 KB - games_v11i1e40438_app2.pdf ]

References


Abbreviations

ACC: anterior cingulate cortex
ADHD: attention deficit/hyperactivity disorder
BEG: bicycle exercise group
CBF: cerebral blood flow
EEG: electroencephalography
ERP: event-related potential
EXG: exergaming group
FAIR: Frankfurter Aufmerksamkeits-Inventar (Frankfurt Attention Inventory)
HR: heart rate
RT: response time
Abstract

Background: Exergaming may be an important option to support an active lifestyle, especially during pandemics.

Objective: Our objectives were (1) to explore whether change in exergaming status (stopped, started or sustained exergaming, or never exergamed) from before to during the COVID-19 pandemic was related to changes in walking, moderate-to-vigorous physical activity (MVPA) or meeting MVPA guidelines and (2) to describe changes among past-year exergamers in minutes per week exergaming from before to during the pandemic.

Methods: A total of 681 participants (mean age 33.6; SD 0.5 years; n=280, 41% male) from the 22-year Nicotine Dependence in Teens (NDIT) study provided data on walking, MVPA, and exergaming before (2017 to 2020) and during (2021) the COVID-19 pandemic. Physical activity (PA) change scores were described by change in exergaming status.

Results: We found that 62.4% (n=425) of the 681 participants never exergamed, 8.2% (n=56) started exergaming during the pandemic, 19.7% (n=134) stopped exergaming, and 9.7% (n=66) sustained exergaming. Declines were observed in all 3 PA indicators in all 4 exergaming groups. The more salient findings were that (1) participants who started exergaming during COVID-19 reported the highest MVPA levels before and during the pandemic and declined the least (mean –35 minutes/week), (2) sustained exergamers reported the lowest MVPA levels during the pandemic (median 66 minutes/week) and declined the most in MVPA (mean change of –92 minutes/week) and in meeting MVPA guidelines (–23.6%). During the pandemic, starting exergamers reported 85 minutes of exergaming per week and sustained exergamers increased exergaming by a median 60 minutes per week.

Conclusions: Although starting and sustaining exergaming did not appear to help exergamers maintain prepandemic PA levels, exergaming can contribute a substantial proportion of total PA in young adults and may still represent a useful option to promote PA during pandemics.

(JMIR Serious Games 2023;11:e41553) doi:10.2196/41553

KEYWORDS
exergaming; active video games; longitudinal study; COVID-19 pandemic; physical activity; serious game; youth; young adult; health promotion; digital health intervention; exergame
Introduction

Many studies conducted early on during the COVID-19 pandemic suggest widespread decreases in physical activity (PA) related to public health lockdowns and restrictions [1-5]. Specifically, working from home was associated with declines in active transport, and the closure of gyms and recreational centers limited access to and opportunities for PA [1-5]. Conversely, some pandemic-related changes may have facilitated PA [6]. With less time needed for commuting to and from work, leisure time may have increased for some, thus affording more time to engage in PA [6].

A review on PA during COVID-19 suggested that “positive technology” (defined as a “scientific and applied approach to the use of technology for improving the quality of our personal experience”) [7,8] represented a popular option to support or maintain an active lifestyle during the pandemic [9,10]. Use of technology has been suggested by the World Health Organization (WHO) as one alternative to stay physically active during confinements [11], and many appear to have turned to positive technology during the COVID-19 outbreak to maintain or increase PA [7,9].

Exergaming (or active video gaming), one form of positive technology that engages users in PA, can be played using gaming consoles, mobile devices, virtual reality, personal computers, and specific types of exercise equipment (eg, stationary bikes and treadmills with interactive screens) [7,10,12]. It offers a PA option that is both enjoyable and motivating because of self-monitoring options, gamification, and user interface technology [7,13]. Pokémon Go, for example, is a notably popular exergame that increased PA in both men and women of all ages and weights, even reaching low-activity populations [12]. Exergaming offers personalized, convenient activity options, as well as the possibility of playing with other users online, which, in addition to decreasing sedentary time, may decrease feelings of social isolation during pandemic-related restrictions [13-15]. Exergaming can help people cope with mental health issues such as anxiety [9,16,17], and it can support those with chronic disease, such as Parkinson diseases [17,18], poststroke sequelae [19], arthritis [20], and brain injuries [21], in maintaining PA levels in home-based rehabilitation situations during pandemics.

Before the COVID-19 pandemic, the prevalence of weekly exergaming among adolescents and young adults ranged from 18% to 43% in Canada [22-25]. However, it is not clear whether the prevalence changed during the COVID-19 pandemic. Further, no longitudinal studies have assessed whether exergaming was associated with changes in PA levels during the pandemic. Because future pandemics are possible, understanding exergaming behavior during the COVID-19 pandemic could shed light on potential benefits and challenges of using technology to encourage PA during lockdowns. Further, stressful times can encourage positive behavior change (eg, taking up exergaming) that may endure postpandemic [13,17,26]. Therefore, building the evidence base in this realm could inform public health decisions related to promoting PA during a pandemic.

The objectives of this study were to (1) explore whether a change in exergaming status (stopped, started or sustained exergaming, or never exergamed) from before to during the COVID-19 pandemic related to changes in walking, minutes of moderate-to-vigorous physical activity (MVPA), and meeting MVPA guidelines. We were particularly interested in ascertaining whether starting to exergame or sustaining exergaming were associated with maintaining pre–COVID-19 PA levels of walking, MVPA, or meeting MVPA guidelines. We also aimed to (2) describe changes in minutes of exergaming per week among past-year exergamers from before to during the COVID-19 pandemic.

Methods

Study Sample

Data were drawn from the Nicotine Dependence in Teens (NDIT) study [27], an ongoing 23-year longitudinal study that aims to describe the natural course of nicotine dependence in youth, but also collects data on a wide range of sociodemographic, substance use, psychosocial, lifestyle, and health-related variables. A detailed description of the NDIT study methods has been published previously [27]. Relevant to this study, data for 799 participants were collected in cycle 23 between January 2017 to March 2020 (n=551, 69% of participants completed the questionnaire in 2017; n=152, 19% in 2018; n=88, 11% in 2019; and n=8, 0.8% in 2020). Data were available for 722 participants in cycle 24, which was conducted online during the COVID-19 pandemic, from December 2020 to February 2021. Participants could use their mobile phone, tablet, or personal computer to complete the questionnaire. A total of 681 participants in cycle 23 (85%) also provided data in cycle 24 and were retained for this analysis. Participants received a CAD $50 (US $36.97) gift card or e-transfer in each cycle to cover any costs associated with their participation.

Ethical Approval

The study procedures were approved by the Montreal Department of Public Health Ethics Review Committee (2007–2384), the McGill University Faculty of Medicine Institutional Review Board (2017–6895), and the Ethics Research Committee of the Centre de Recherche du Centre Hospitalier de l’Université de Montréal (2021-9385, 20.278-YP). Parental consent was obtained at NDIT inception. Participants could legally provide consent in the post–high school data collections because they had all attained the age of 18 years.

Study Variables

PA Variables

Data on number of minutes per week walking, moderate physical activity, and vigorous physical activity were collected using the short form of the International Physical Activity Questionnaire (IPAQ-SF) [28,29]. The IPAQ-SF is used in cross-national monitoring of PA and demonstrates low to moderate reliability and validity against device-based PA measures [30]. Participants reported the number of days on which they had engaged in each type of PA over the past week and the average number of minutes per bout. The recommended truncation protocol [29] was used to exclude unrealistic values and reduce improbable...
PA scores. Number of minutes walking per week was calculated as the number of days walking per week multiplied by the number of minutes per bout. Number of moderate PA minutes per week was computed as the number of days of moderate PA per week multiplied by the number of minutes per bout, and the number of vigorous PA minutes per week was computed as the number of days of vigorous PA per week multiplied by the number of minutes per bout.

**Calculation of MVPA**

Number of MVPA minutes per week was calculated by adding the number of moderate PA minutes per week and number of vigorous PA minutes per week.

**Meeting MVPA Guidelines**

Participants were categorized as meeting MVPA guidelines (yes or no) if they had engaged in MVPA for ≥150 minutes per week [31].

**Exergaming Status**

Using data from cycles 23 and 24, participants were categorized into 4 groups based on their exergaming behavior over time using the following item: “In the past 12 months, how often did you exergame using consoles, or using your cell phone and/or a mobile app?” Response choices included never, less than once a month, 1 to 3 times a month, 1 to 6 times per week, and every day. In both cycles 23 and 24, participants were categorized as exergamers if they chose any response option other than “never.” The four groups included (1) never-exergamers (ie, participants who did not report exergaming in cycles 23 or 24); (2) stopping exergamers (ie, participants who reported exergaming in cycle 23 but not in cycle 24); (3) starting exergamers (ie, participants who reported exergaming in cycle 24 but not in cycle 23); and (4) sustaining exergamers (ie, participants who exergamed in both cycles 23 and 24).

**Minutes Exergaming per Week**

Items measuring the number of minutes exergaming per week were modeled on the IPAQ-SF [22,23,25]. Participants who reported past-year exergaming were asked: “Did you exergame using consoles such as Nintendo Wii, XBOX ONE Kinect, Sony Play Station Move, Sony Eye Toy: Kinetic, or using your cell phone and/or a mobile app?” Response choices included never, less than once a month, 1 to 3 times a month, 1 to 6 times per week, and every day. In both cycles 23 and 24, participants were categorized as exergamers if they chose any response option other than “never.” The four groups included (1) never-exergamers (ie, participants who did not report exergaming in cycles 23 or 24); (2) stopping exergamers (ie, participants who reported exergaming in cycle 23 but not in cycle 24); (3) starting exergamers (ie, participants who reported exergaming in cycle 24 but not in cycle 23); and (4) sustaining exergamers (ie, participants who exergamed in both cycles 23 and 24).

**Exergaming Intensity**

Usual intensity of exergaming was measured by the following question: “What was your usual physical effort during play?” Response choices included light, moderate, and vigorous.

**Sociodemographic Characteristics**

Sociodemographic characteristics included age, sex, whether the participant was university educated (yes or no), whether their mother was university educated (yes or no), and whether they were French speaking (yes or no), were born in Canada (yes or no), had an annual household income <CAN $50,000 (US $36,966; yes or no), lived alone (yes or no), lived with children (yes or no), and were employed (yes or no).

**Data Analyses**

Descriptive statistics were used to compare sociodemographic and PA-related characteristics (ie, walking, MVPA, and exergaming) of participants retained in the analytic sample (ie, those who provided data in cycles 23 and 24) versus not retained (ie, those who provided data in cycle 23 but not cycle 24). Based on skewness and kurtosis, walking, MVPA, and exergaming minutes per week were not normally distributed. Therefore, the median (IQR) is reported for these variables. The proportion of participants that met MVPA guidelines is reported as a percentage.

Change scores were computed by subtracting number of minutes per week for each PA indicator (including exergaming) in cycle 23 from minutes per week in cycle 24. Because the distributions for the change scores were relatively normal in each of the 4 groups defined by exergaming status as well as in the sample overall, we report mean (SD) change scores. We report the median (IQR) for the change in minutes exergaming per week because the data were not normally distributed. Data were analyzed using SPSS (version 20.0; IBM Corp).

**Results**

**Participants Retained Versus Not Retained**

Table 1 compares the sociodemographic and PA-related characteristics of participants retained and not retained (ie, as measured in cycle 23). Lower proportions of those retained were male and French speaking. In addition, they reported fewer minutes per week walking and engaging in MVPA than participants who were not retained.

The mean age of the 681 participants retained for analysis was 33.6 (SD 0.5) years in cycle 24; 41% (n=280) were male, 95% (n=647) were born in Canada, 33% (n=225) were French speaking, 21% (n=143) had an annual household income <CAN $50,000 (US $36,966), 15% (n=102) lived alone, 51% (n=347) lived with children, 83% (n=565) reported being employed, and 36% (n=245) met MVPA guidelines. The median (IQR) minutes per week during COVID-19 was 120 (45-240) for walking and 80 (0-240) for MVPA. Participants who exergamed during COVID-19 (n=122) reported a median (IQR) of 80 (9-240) minutes of exergaming per week.

Of 681 participants retained for analysis, 425 (62.4%) were never-exergamers, 56 (8.2%) started exergaming in cycle 24 (during the pandemic), 134 (19.7%) stopped exergaming in cycle 24, and 66 (9.7%) sustained exergaming. Among the 200 exergamers in cycle 23, 66 (33%) sustained exergaming in cycle 24, and 134 (67%) stopped exergaming.
Table 2 reports the median number of minutes walking per week, median number of minutes engaged in MVPA per week, and the percentage of participants who met MVPA guidelines before and during the COVID-19 pandemic. In addition to reporting values for the sample overall, data are provided for the 4 groups defined by exergaming behavior over time.

The data suggest declines over time in all 3 PA indicators in the sample overall (Table 2). Median minutes walking per week declined from 175 before COVID-19 to 120 during COVID-19, with a mean decline of 58 minutes per week. MVPA declined from a median 130 to 80 minutes per week among the participants overall, with a mean decline of 49 minutes per week. The proportion of participants that met MVPA guidelines declined from 48.3% (n=326) before the pandemic to 36% (n=245) during the pandemic, an absolute decline of 12.3%.

Table 1. Characteristics of participants retained and not retained for analysis from the Nicotine Dependence in Teens study (Montreal, Quebec, 2017-2021). Data were drawn from cycle 23a.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (n=799)</th>
<th>Retained in cycle 24 (n=681)</th>
<th>Not retained in cycle 24 (n=118)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>30.6 (1.0)</td>
<td>30.5 (1.0)</td>
<td>31.0 (1.1)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>351 (43.9)</td>
<td>280 (41.1)</td>
<td>71 (60.2)</td>
</tr>
<tr>
<td>Born in Canada, n (%)</td>
<td>799 (93.7)</td>
<td>642 (94.3)</td>
<td>107 (90.7)</td>
</tr>
<tr>
<td>Mother is university educated, n (%)</td>
<td>332 (46.2)</td>
<td>283 (45.6)</td>
<td>49 (50)</td>
</tr>
<tr>
<td>French speaking, n (%)</td>
<td>245 (30.7)</td>
<td>199 (29.2)</td>
<td>46 (39)</td>
</tr>
<tr>
<td>Walking (minutes/week), median (IQR)</td>
<td>180 (20-450)</td>
<td>135 (0-303)</td>
<td>135 (0-303)</td>
</tr>
<tr>
<td>MVPAc (minutes/week), median (IQR)</td>
<td>135 (0-330)</td>
<td>135 (0-330)</td>
<td>180 (20-450)</td>
</tr>
<tr>
<td>Met MVPA guidelines, n (%)</td>
<td>387 (49.5)</td>
<td>326 (48.9)</td>
<td>61 (53)</td>
</tr>
<tr>
<td>Exergaming (minutes/week), median (IQR)c</td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td>0 (0-83)</td>
</tr>
</tbody>
</table>

aTotals may differ due to missing data.
bIncludes only past-year exergamers.
cMVPA: moderate-to-vigorous physical activity.
Table 2. Number of minutes walking per week, number of minutes engaged in moderate-to-vigorous physical activity per week, and percentage of participants who met moderate-to-vigorous physical activity guidelines from before to during the COVID-19 pandemic in 4 groups defined by consistency in exergaming behavior over time (Nicotine Dependence in Teens study, Montreal, Canada, 2017-2021).a,b,c,d,e,f

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (n=681)</td>
<td>175 (60-350)</td>
<td>120 (45-240)</td>
<td>−58 (310)d</td>
</tr>
<tr>
<td>Never exergamed (n=424)</td>
<td>175 (60-315)</td>
<td>120 (45-240)</td>
<td>−48 (298)d</td>
</tr>
<tr>
<td>Started exergaming (n=56)</td>
<td>163 (90-420)</td>
<td>195 (40-420)</td>
<td>−38 (372)d</td>
</tr>
<tr>
<td>Stopped exergaming (n=134)</td>
<td>178 (88-420)</td>
<td>120 (60-210)</td>
<td>−97 (279)d</td>
</tr>
<tr>
<td>Sustained exergaming (n=66)</td>
<td>140 (30-258)</td>
<td>90 (40-188)</td>
<td>−56 (380)d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MVPA (minutes/week), median (IQR)</th>
<th>Overall (n=681)</th>
<th>Never exergamed (n=424)</th>
<th>Started exergaming (n=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (n=681)</td>
<td>130 (0-310)</td>
<td>120 (0-300)</td>
<td>210 (40-420)</td>
</tr>
<tr>
<td>Never exergamed (n=424)</td>
<td>80 (0-240)</td>
<td>80 (0-240)</td>
<td>90 (4-394)</td>
</tr>
<tr>
<td>Started exergaming (n=56)</td>
<td>180 (41-360)</td>
<td>80 (0-275)</td>
<td>66 (0-221)</td>
</tr>
<tr>
<td>Stopped exergaming (n=134)</td>
<td>180 (0-360)</td>
<td>66 (0-221)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Met MVPA guidelines, n (%)</th>
<th>Overall (n=681)</th>
<th>Never exergamed (n=424)</th>
<th>Started exergaming (n=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (n=681)</td>
<td>326 (48.3)</td>
<td>186 (44.8)</td>
<td>31 (57.1)</td>
</tr>
<tr>
<td>Never exergamed (n=424)</td>
<td>245 (36)</td>
<td>155 (36.6)</td>
<td>21 (37.5)</td>
</tr>
<tr>
<td>Started exergaming (n=56)</td>
<td>74 (55.6)</td>
<td>48 (35.8)</td>
<td>21 (31.8)</td>
</tr>
<tr>
<td>Sustained exergaming (n=66)</td>
<td>35 (55.4)</td>
<td>21 (31.8)</td>
<td></td>
</tr>
</tbody>
</table>

aTotals may differ due to missing data.
bData were collected from January 2017 to March 2020 (mean age 30.6 years).
cData were collected from December 2020 to June 2021 (mean age 33.6 years).
dThese values are the mean (SD), since the change scores were normally distributed.
eMVPA: moderate-to-vigorous physical activity.
fThese values represent the absolute decline in percentage.

Change in Walking Across Exergaming Groups

Among the 4 groups defined by exergaming status, patterns of interest in walking from before to during COVID-19 (Table 2) show that those who started exergaming reported the highest level of walking during the pandemic (a median 195 minutes per week), and in addition, they declined the least (they had a mean decline of 38 minutes per week, compared to 58 minutes among all participants). Those who stopped exergaming had the highest level of walking before COVID-19 (median 178 minutes per week) and declined the most (their mean minutes walking per week declined by 97 minutes). Finally, sustained exergamers reported the lowest median levels of walking both before and during the pandemic; the mean decline in minutes walking per week in this group (−56 minutes per week) was similar to the mean decline among participants overall (−58 minutes per week).

Change in MVPA Across Exergaming Groups

Patterns of interest in MVPA in the 4 exergaming status groups (Table 2) show that never-exergamers reported the lowest MVPA levels before the COVID-19 pandemic (a median 120 minutes per week). Participants who started exergaming during the pandemic reported the highest MVPA levels both before and during the COVID-19 pandemic, and they declined the least (Table 2). Their mean MVPA change score was −35 minutes MVPA per week compared to −49 minutes among participants overall. Finally, sustained exergamers reported the lowest MVPA levels during the pandemic (a median 66 minutes per week) and they declined the most. Their mean MVPA change score was −92 minutes per week.

Change in Meeting MVPA Guidelines Across Exergaming Groups

Compared to the other 3 groups, a higher proportion of participants who started to exergame during the COVID-19...
pandemic met MVPA guidelines both before and during the COVID-19 pandemic. Although the proportion of sustained exergamers that met MVPA guidelines was similar to the other 2 exergaming groups before COVID-19, this group reported the lowest proportion during COVID-19 and the largest decline over time (−23.6%). The proportion of participants that met MVPA guidelines was lowest among never-exergamers before COVID-19 (186/424, 44.8%), and this group had the lowest decline over time (−8.2%).

Change in Minutes Exergaming Among Exergamers

Table 3 reports change in exergaming minutes per week among sustained exergamers, participants who started exergaming, and participants who stopped exergaming. Participants who started exergaming reported a median of 85 minutes of exergaming per week during the COVID-19 pandemic. Sustained exergamers also increased exergaming by a mean 90.5 minutes per week (Table 3). Light intensity exergaming was reported by 65% (47/72) of exergamers in cycle 23 and 59% (55/94) of exergamers in cycle 24. Moderate or vigorous exergaming was reported by 35% (25/72) of exergamers in cycle 23 and 42% (39/94) in cycle 24.

Table 3. Number of minutes per week engaged in exergaming before and during the COVID-19 pandemic in groups defined by consistency in exergaming behavior over time (Nicotine Dependence in Teens study, Montreal, Quebec, 2017-2021).

<table>
<thead>
<tr>
<th>Exergaming (minutes/week), median (IQR)</th>
<th>Before COVID-19 (cycle 23)</th>
<th>During COVID-19 (cycle 24)</th>
<th>Change between cycles 23 and 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Started exergaming (n=56)</td>
<td>0 (0 to 0)</td>
<td>85 (0 to 240)</td>
<td>85 (0 to 240)</td>
</tr>
<tr>
<td>Stopped exergaming (n=134)</td>
<td>0 (0 to 0)</td>
<td>0 (0 to 0)</td>
<td>0 (0 to 0)</td>
</tr>
<tr>
<td>Sustained exergaming (n=66)</td>
<td>0 (0 to 63)</td>
<td>68 (14 to 229)</td>
<td>60 (−4 to 188)</td>
</tr>
</tbody>
</table>

aData were collected from January 2017 to March 2020 (mean age 30.6 years).
bData were collected from December 2020 to June 2021 (mean age 33.6 years).
cParticipants who reported past-year exergaming but 0 minutes exergaming in the past 30 days were assigned a score of 0 minutes exergaming per week.

Discussion

Overall Findings

Overall, the percentage of participants meeting MVPA guidelines, the number of minutes walked, and the number of minutes engaged in MVPA per week declined from before to during the pandemic regardless of exergaming status [6]. Our data suggest that sustained exergamers were not more active than never-exergamers during the COVID-19 pandemic and, in fact, appeared to be less active. In contrast, those who started exergaming reported the highest levels of walking and MVPA during COVID-19, although, as in all groups, the change scores suggest declines on average in both activities. Thus, it appears that even if exergaming encourages PA during periods of confinement, in and of itself exergaming may not be sufficient to help exergamers maintain pre-pandemic PA levels. It is possible, however, that without exergaming the declines observed among sustained exergamers could have been steeper.

Never-Exergamers

Never-exergamers reported declines of 48 minutes per week in walking and 41 minutes per week in MVPA from before to during the COVID-19 pandemic. Of the 4 exergaming groups studied, never-exergamers reported the lowest levels of MVPA and were least likely to meet MVPA guidelines before the pandemic. The relatively minor changes in PA in this group during COVID-19 may reflect a general disinterest in PA (so that closure of gyms and recreational centers during the pandemic made little difference to their PA levels). It is possible that level of motivation for PA remained low among these individuals during the pandemic, such that they did not benefit from any increases in time available for PA due to containment measures. It is also possible that this group had established (lower) PA levels and patterns that they were content with and that they felt little need to explore different ways of building PA into their routines, regardless of the pandemic context.

Stopped Exergaming

Among the 4 groups defined by exergaming status, those who stopped exergaming declined the most in walking (–97 minutes/week) from before to during the COVID-19 pandemic. Although stopping exergaming in this group was possibly related to COVID-19, the middle to early 30s is often a turbulent time during the life course marked by numerous important transitions as people complete their education, enter the workforce, and begin their own families [32]. Rather than being a consequence of the COVID-19 pandemic, exergaming may be a form of PA that is dropped during these life transitions to make way for engaging in new roles and activities. Alternatively, exergaming could be an intermittent activity [32] linked to the release of new games or consoles, and the lack of new releases during the pandemic could be among the reasons that participants stopped exergaming. Exergaming may also be a “transferable” PA, such that it is used to experiment with a new PA (eg, to begin jogging), which is then continued independently of the exergaming component. Finally, frustration with technical glitches, concerns with online privacy [33], reduction in enjoyment and game immersion (ie, how a videogame draws a player into the game) [33,34], and price have also been cited as reasons for stopping exergaming [33,34]. Qualitative studies may be helpful in identifying the reasons why exergamers in their 30s choose to stop exergaming during pandemics.
Sustained Exergaming

Of the 4 exergaming groups, sustained exergamers reported the lowest levels of walking both before and during the pandemic and the lowest levels of MVPA during the pandemic. Previous research suggests that although exergamers do not necessarily have higher PA levels than never-exergamers and may not be interested in traditional PA, they do understand the importance of movement and turn to exergaming for their PA [22,25,35]. Sustained exergamers may have continued to play games such as Pokémon Go during the COVID-19 pandemic. There is evidence that 2020 was the most profitable year ever for the Pokémon Go enterprise [36], suggesting that many exergamers continued to play this outdoor game during COVID-19. Zombies, Run! also had an increase of 2 million users during COVID-19 [37]. Further, Ellis et al [38] reported that among 2004 young adult gamers (aged 30.5 years on average), Pokémon Go and Harry Potter: Wizards Unite were played frequently during the pandemic to maintain exercise levels and for social connection. Many participants reported that Pokémon Go was “the one thing that keeps me going outdoors and moving each day” [34]. Ellis et al [38] also reported that those who exergamed during the COVID-19 pandemic did so to distract themselves, escape from reality for a short period of time, occupy themselves, and manage their mental health.

Sustained exergamers reported a median of 68 minutes per week of exergaming, an important contribution to overall PA. Exergaming is not usually dependent on fitness or recreational facilities outside the home; therefore, this group may not have been overly affected by pandemic-related lockdowns. While the change in walking levels among sustained exergamers was not as marked as among those who stopped exergaming, their decline in MVPA was substantial (~92 minutes per week). It is possible that sustained exergamers participated in other modes of PA before the pandemic that were not replaced during the pandemic. Minutes exergaming per week among sustained exergamers did increase (as also reported in the Ellis et al study), but this was not enough to replace MVPA lost in other PA modes.

Started Exergaming

Of the 4 exergaming groups, those who started exergaming during the pandemic reported the highest MVPA levels before the pandemic (210 minutes/week) and the smallest decrease in MVPA (~35 minutes/week) during the pandemic. This group may have started using exergaming to compensate for activities they could no longer participate in during COVID-19. In addition to compensating for the loss of PA during the pandemic, video games may have helped with the loss of social interactions, which might have been a particularly salient loss in this group [7,14,38,39]. Whatever the underpinning, starting exergaming during the pandemic likely reflects relatively high levels of “physical literacy,” which represents the motivation and financial resources to search out new forms of accessible PA [40]. Alternatively, or in addition, this group may also have been particularly sensitive to public health messaging that recommended maintaining PA levels during the pandemic.

Future Research

More research is needed to examine fluctuations in exergaming behaviors over time, especially during such challenging times as COVID-19. Qualitative research may help identify reasons for stopping, starting, or continuing to exergame during pandemics. This study did not identify types of exergaming engaged in, who participants exergamed with, or the context (ie, indoors or outdoors) in which they exergamed. These areas could be important avenues to explore to better inform recommendations for starting exergaming or for maintaining previous levels of exergaming during pandemics, as well as in general.

Limitations

Limitations of this study include that self-report IPAQ-SF data are subject to overreporting, although Lee et al [28] suggest that because of its established reliability, the IPAQ-SF can be used with care in repeated-measures studies. There was loss to follow-up between cycle 23 and cycle 24, which may relate to competing interests in the busy lives of young people in emerging adulthood. Selection bias related to this loss to follow-up could have attenuated our estimates, although the small number of sociodemographic differences between those retained and not retained for analysis mitigates this concern somewhat. Use of a purposive sample of schools in the NDIT study may limit the generalizability of the findings. Finally, we were unable to categorize exergamers according to more refined categories of exergaming intensity and frequency, and this may have led to misclassification of some exergamers. Future studies should use more refined measures of exergaming to enable more accurate categorization of exergamers.

Conclusion

Although it may support PA at home during periods of confinement, starting or sustaining exergaming did not appear to be enough to maintain prepandemic PA levels in this population-based sample of young adults. However, the data suggest that exergaming can contribute a substantial proportion of total PA in young adults and may still represent a useful option to promote PA during pandemics.

Acknowledgments

The Nicotine Dependence in Teens study was supported by the Canadian Cancer Society (grants 010271, 017435, and 704031) and the Canadian Institutes of Health Research (grant 451832). RHO holds a Sciences and Humanities Research Council of Canada (SSHRC) Canada Graduate Scholarships—Doctoral Program (CGS-D) (#767-2022-0015). The funders were not involved in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.
Authors' Contributions

EKO is the first author of the manuscript and conducted all analyses and data verifications. EKO, CMS, RHO, MB, MPS, and JLO drafted and revised the manuscript. JLO designed and acquired the funding for the Nicotine Dependence in Teens study, and JLO, MPS, and MB acquired the funding for the COVID-19 data collection. All authors contributed to the interpretation of results and revised and approved the final manuscript. EKO holds a postdoctoral salary award from the Fonds de Recherche du Québec-Santé (FRQ-S). CS holds a Canada Research Chair in physical activity and mental health. MPS holds a J2 Salary award from the FRQ-S. JLO held a Canada Research Chair in physical activity in the Early Determinants of Adult Chronic Disease from 2004 to 2021.

Conflicts of Interest

None declared.

References


O’Loughlin E. Contribution of exergaming behaviour to physical activity: Toward better understanding the role of motivation.. Concordia University. 2019. URL: https://spectrum.library.concordia.ca/id/eprint/986369/1/oloughlin_PhD_S2020.pdf [accessed 2022-02-14]


O’Loughlin EK, Barnett TA, McGrath JJ, Conshalvo M, Kakinami L. Factors associated with sustained exergaming: Longitudinal investigation. JMIR Serious Games 2019 Jul 31;7(2):e13335 [FREE Full text] [doi: 10.2196/13335] [Medline: 31368440]


Abbreviations

IPAQ-SF: International Physical Activity Questionnaire–Short Form
MVPA: moderate-to-vigorous physical activity
NDIT: Nicotine Dependence in Teens
PA: physical activity
WHO: World Health Organization
Usability of the GAIMplank Video Game Controller for People With Mobility Impairments: Observational Study

Laurie A Malone¹, MPH, PhD; Christen J Mendonca¹, PhD; Sangeetha Mohanraj¹, MS; Samuel R Misko², MSEE; Joseph Moore², BSEE; James Michael Brascome², MS; Mohanraj Thirumalai¹, PhD

¹School of Health Professions, The University of Alabama at Birmingham, Birmingham, AL, United States
²Engineering and Innovative Technology Development, The University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Laurie A Malone, MPH, PhD
School of Health Professions
The University of Alabama at Birmingham
3810 Ridgeway Dr.
Birmingham, AL, 35209
United States
Phone: 1 205 934 0142
Email: lamalone@uab.edu

Abstract

Background: Replacing sedentary behaviors during leisure time with active video gaming has been shown to be an enjoyable option for increasing physical activity. However, most off-the-shelf active video gaming controllers are not accessible or usable for individuals with mobility impairments. To address this requirement, a universal video game controller (called the GAIMplank) was designed and developed.

Objective: This study aimed to assess the usability of the GAIMplank video game controller for playing PC video games among individuals with mobility impairments. Measures of enjoyment, perceived exertion, and qualitative data on the user experience were also examined.

Methods: Adults (aged 18-75 years) with a mobility impairment were recruited to participate in a single testing session in the laboratory. Before testing began, basic demographic information, along with minutes of weekday and weekend physical activity, minutes of weekday and weekend video game play, and video game play experience were collected. The GAIMplank was mapped to operate as a typical joystick controller. Depending on their comfort and functional ability, participants chose to play seated in a chair, standing, or in their own manual wheelchair. Leaning movements of the trunk created corresponding action in the game (ie, lean right to move right). The participants played a total of 5 preselected video games for approximately 5 minutes each. Data were collected to assess the usability of the GAIMplank, along with self-efficacy regarding execution of game play actions, rating of perceived exertion and enjoyment for each game, and overall qualitative feedback.

Results: A total of 21 adults (n=15, 71% men; n=6, 29% women) completed the usability testing, with a mean age of 48.8 (SD 13.8; range 21-73) years. Overall, 38% (8/21) of adults played while standing, 33% (7/21) of adults played while seated in a chair, and 29% (6/21) played in their own manual wheelchair. Scores from the System Usability Scale indicated above average (74.8, SD 14.5) usability, with scores best for those who played seated in a chair, followed by those standing, and then individuals who played seated in their own wheelchairs. Inconsistencies in the responsiveness of the controller and general feedback for minor improvements were documented. Rating of perceived exertion scores ranged from light to moderate intensity, with the highest scores for those who played seated in a chair. Participants rated their experience with playing each game from above average to very enjoyable.

Conclusions: The GAIMplank video game controller was found to be usable and accessible, providing an enjoyable option for light-to-moderate intensity exercise among adults with mobility impairments. Minor issues with inconsistencies in controller responsiveness were also recorded. Following further development and refinement, the next phase will include a pilot exercise intervention using the GAIMplank system.

(JMIR Serious Games 2023;11:e38484) doi:10.2196/38484
KEYWORDS
exergaming; physical disability; equipment design; video gaming; physical activity; exercise; wheelchair

Introduction

Background
Physical activity is an important component of a healthy lifestyle. A myriad of health benefits, both physical and psychological, can be achieved with daily physical activity [1]. Physical activity options, however, are limited for people with mobility impairments. Issues with accessibility and transportation are experienced by many individuals wanting to fulfill their daily physical activity needs. Fitness center parking, inaccessible entrances, limited staff knowledge on inclusive programming, and cost of membership and transportation are some of the limiting factors [2]. Equipment such as a stationary arm ergometer can provide a means to exercise at home but may be seen as boring and difficult to adhere to by many [3]. In an effort to combat the tedium of routine exercise, active video gaming (AVG) has become popular.

Replacing sedentary behaviors during leisure time with AVG has been shown to be an enjoyable option for increasing the amount of physical activity acquired each week [4-12]. The promise of AVGs is particularly appealing for people with disabilities, given the high rate of inaccessible features in the built environment and the difficulty in finding fun and engaging activities [5-8]. Because AVG systems are relatively affordable, they also hold promise as a scalable product for promoting improved health outcomes and higher levels of physical activity and fitness among people with disabilities [13]. Several AVG systems have been in the market for the last 2 decades; however, most are not adapted to suit the needs of individuals with mobility impairments. Using a mixed methods study, the usability of a somatosensory square dance system for older adults was examined [14]. The system was developed to support a popular fitness activity and to eliminate some of the common barriers (risk of injury, noise, and space) experienced by older adults. The system was well received by the participants, providing them with a fun fitness activity that could be safely performed indoors in a private space. The participants included a group of older adults (aged 55-68 years); however, specific adaptation for individuals with mobility impairments was not the focus and therefore was not considered. Another study examined the usability of an AVG platform for individuals with mobility impairments. After several rounds of usability testing with residents, transport, and family, the feedback indicated that the system was an enjoyable way to engage in physical activity. Although efforts are being made, there continues to be a pressing need to make AVG controllers accessible to people who are unable to stand for long periods, cannot stand on a small platform because of poor balance or extreme obesity, or those who use a wheelchair for all daily activities.

Our engineering design and development team, with the Rehabilitation Engineering Research Center on Interactive Exercise Technologies and Exercise Physiology for People with Disabilities, previously built a proof-of-concept (PoC) wheelchair-accessible adapted gaming controller for the Wii system [16] and our research team tested its usability rating compared with the commercial off-the-shelf (OTS) balance board controller provided with the product [17]. Participants with mobility impairments found the wheelchair-accessible version to be more usable compared with the Wii OTS board (P<.01). The mean usability scores were 71.7 (SD 18.03) for the adapted controller and 32.1 (SD 36.72) for the Wii OTS board (higher scores reflect greater usability). In addition, a significant negative correlation (r=0.692; P<.001) was found between lower extremity function and system usability scores, indicating that for participants with poorer lower extremity function, the adapted board was perceived as more usable. Furthermore, participants reported activity using the adapted board to be enjoyable, achieving light-to-moderate intensity exercise [18].

As a next step, the aim was to transition from a PoC gaming controller to the performance of highly targeted proof-of-product (PoP) activities for the design of a new wheelchair-accessible AVG controller using the expertise and input of engineers, product designers, potential users, and other stakeholders from the community. The goal was to convert from a Wii-only interface to a universal controller for use with all games available to play on a PC.

Study Objectives
Following the design and development process, this study aimed to assess the usability of a wheelchair-accessible GAimplank video game controller for individuals with mobility impairments. Measures of enjoyment, perceived exertion, and qualitative data on user experience were also collected.

Methods

PoP Development Process
To begin our PoP development effort, approximately 30 semistructured interviews were conducted with end users in various market segments to discover use cases that would be used to drive design features, functions, and capabilities. The physical therapy clinic market segment was found to be lacking in financial resources for investment into technology for nonreimbursable activities such as AVG. The market segment for inpatient rehabilitation and retirement facilities and communities was found to have the financial resources for recreational activities but lacked robust end user interest. The universal video game accessory market was found to be much more robust in terms of interest for nontraditional game controllers to facilitate intuitive player movement in virtual reality environments as well as to encourage AVG. Effective entry into this market requires the game controller to (1) be easily transportable and stowable, (2) have a purchase price comparable with other high-end game controllers (<US $300), and (3) be universally compatible with all major gaming platforms.
The design process began based on the PoC prototype developed previously (Figure 1), which used a custom-built aluminum frame, user interface, and load cell sensors combined with repurposed internal electronics from a deconstructed Wii Fit Board to allow for wireless connection to Wii game consoles. On the basis of the market study results, the preliminary PoP design effort targeted 3 areas for improvement, including reducing the weight of the balance board to <6.8 kg, integrating custom electronics for universal compatibility with all major gaming platforms, and optimizing design for low-cost manufacturing.

A redesign of the PoC balance board platform began with the determination of the critical dimensions for use by standing and seated users while also minimizing the footprint and structural support of the weightbearing structure. Early conceptual designs featured novel solutions to these design challenges through the integration of handrails, footrests, and ramps into the platform design (Figure 2). After evaluating a series of finite element models and simplified prototypes, a resin-infused sandwich composite was the material selected for use to reduce the weight of the balance board while maintaining the stiffness necessary to redistribute the user’s weight to a minimum number of load cells.

The redesign of the PoC balance board electronics began with the selection of the embedded processor, load cell measurement integrated circuits, and wired and wireless communication modalities. The user interface components such as the sensitivity adjustments (overall, front and back, right and left) were upgraded from the PoC’s buttons and knobs to an Android app that can easily provide access to any number of different controller settings and functions to optimize the user’s game play experience. A Raspberry Pi Zero was implemented as the primary processor facilitating Bluetooth connection to an Android tablet as well as allowing remote access to data and programming via Wi-Fi. The USB Human Interface Device periphery connection to PCs and gaming console adapters was facilitated using an ARM Cortex-M4 microcontroller configured as a generic joystick device. Load cell measurements were performed using dedicated low-noise 24-bit analog-to-digital microchips.

The final design of the GAIMplank balance board platform (Figure 3) features rims on 3 sides of the board that serve as a physical barrier at the edge of the board. A downward tapered edge along the back of the board, and a slimline ramp, allowed for easy roll-on access (eg, if the user is in a wheelchair). The GAIMplank platform also features yellow high-visibility graphics that provide the user with a sense of their alignment on top of the board. The weight capacity of the GAIMplank platform is ≤270 kg, which is sufficient to support a wide range of users.

The final design of the GAIMplank Android App features real-time bidirectional information flow and access to a large range of customizable settings for real-time adjustments (Figure 4) to the measured load cells’ bias, dead zones, and sensitivity, as well as to allow for on-demand calibration of the GAIMplank’s output based on a player’s measured weight and range of movement and game play mode (joystick vs directional pad).

The final design of GAIMplank electronics includes a custom printed circuit board (PCB) stack and an LED matrix display (Figure 5). The need for this display was discovered during the early developmental testing of prototypes with 3 volunteers recruited from local video gaming clubs, whereby they required instant feedback to build confidence that the board was indeed responding accurately and quickly to their movements during game play. This display is also used to guide the user through the required movements during calibration (step off, step on, move forward and backward, and move left and right). The PCB stack includes a Raspberry Pi as the central processor for the device, controlling wireless and wired communication, the LED matrix, and the measurement of the position of the user’s center of balance at a rate of approximately 50 Hz. The stack also provides robust fixturing for the 22-pin D-Sub connector, which is used to connect to the load cells embedded in the platform via a flat flexible cable, as well as for the mini USB port for a wired connection to the gaming PC for the game console adapter.

To ensure safety during game play, height-adjustable handrails were located on 3 sides of the GAIMplank. In addition, a Microsoft adaptive controller was integrated into the system to facilitate the use of external buttons and triggers used during game play.
Figure 1. Proof-of-concept balance board prototype.

Figure 2. Early proof-of-product conceptual design for inpatient rehabilitation market.

Figure 3. Final proof-of-product design and prototype for the GAIMplank video game controller. Photos on the left show the top view (upper left) and bottom view (lower left) of the board before the final coating and graphics were applied (right).
Figure 4. Picture of the GAIMplank app with slider bars used to calibrate the device; select mode of play; and adjust sensitivity, bias, and dead zone area.

Figure 5. Proof-of-product design of the GAIMplank electronics stack.
Usability Testing

Design and Setting

Data collection for usability testing was performed at the RERC RecTech Exercise Science and Technology Laboratory (University of Alabama at Birmingham). For the purposes of this study, participants came to the laboratory for a single visit, which lasted for approximately 60-90 minutes.

Ethics Approval

All study procedures were approved by the institutional review board at the University of Alabama at Birmingham (IRB-300003265).

Participant Recruitment

Participants were recruited via flyers and word of mouth at a local community health and fitness center for individuals with physical disabilities and chronic health conditions. To assess the usability of the GAIMplank for different play styles (sitting in a chair, personal wheelchair, and standing), recruitment was stratified. Sample size estimates were based on identifying common usability barriers and issues specific to the 3 game play styles. According to Cazañas et al [19], a sample of 17 individuals would reasonably identify 80% of common problems in the system, with groups of 4 to 9 sufficient to identify problems specific to the mode of play. Additional participants were recruited to account for modest attrition.

Inclusion criteria for potential participants were (1) age 18-75 years; (2) a self-reported lower extremity mobility disability (eg, spina bifida, cerebral palsy, muscular dystrophy, >1 year after spinal cord injury, multiple sclerosis, stroke, or limb loss) with partial or full use of the upper extremities; (3) mobility impairment (ie, gait deviation and balance issue) or use of an assistive device (manual wheelchair, walker, crutches, and canes) for balance or mobility purposes; (4) ability to use arms for exercise; and (5) body weight <180 kg. Exclusion criteria included (1) significant impairment in visual acuity that prevents seeing a 52” television screen to follow exercise, (2) cardiovascular disease event within the past 6 months, (3) severe pulmonary disease or renal failure, (4) current pregnancy, (5) ongoing exacerbation of a health condition, and (6) other conditions that would interfere with intervention or testing procedures. Following the distribution of a flyer, the project recruitment coordinator answered calls from or met interested individuals. If contacted, the recruitment coordinator reviewed the inclusion and exclusion criteria using a screening form to determine whether they were eligible to participate.

Testing Session

Upon arrival at the laboratory, all study procedures were reviewed with the participant. After being given an opportunity to ask questions, the participants provided informed consent. Before the testing began, basic demographic information, along with information regarding minutes of weekday and weekend physical activity, minutes of weekday and weekend video game play, and video game play experience were collected. All procedures for data collection and video game operation were reviewed with the participants before game play. The participants were then settled at the GAIMplank station. Before playing the first game, a calibration procedure was conducted by following a pattern of movements as indicated on the LED matrix.

The GAIMplank was mapped to operate as a typical joystick controller. Leaning movements of the trunk created corresponding actions in the game (ie, lean forward to move character forward and up, lean right to move character right; Figure 6). The participants played a total of 5 preselected video games. The game Feather, a slow-paced flying game, was played first by all participants to introduce them to the movements required for operating the GAIMplank. The order in which the remaining 4 games were played was randomized. To minimize offensive content, the games had an Entertainment Software Rating Board score of E (Everyone), E+10, or Teen. The description of each game played and movements required is provided in Table 1. Each game was played for approximately 5 minutes.
Figure 6. Participants standing (right), seated in a chair (upper left), or wheelchair (bottom left) engaging in active video gaming using the GAIMplank controller. Game play movement was controlled by shifting the body weight or leaning forward, back, right, left as needed. External buttons to activate actions such as jump or shoot could be configured in various formations on a flexible arm within the player’s reach. A trigger button used for acceleration or shooting could be held in the hand or placed on the external button pad.

Table 1. Description of games played and movements required.

<table>
<thead>
<tr>
<th>Game</th>
<th>Genre</th>
<th>Game description</th>
<th>Trunk leaning movements required</th>
<th>External buttons used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feather</td>
<td>Flight; exploration</td>
<td>Relaxing flight to explore a scenic nature landscape</td>
<td>Left, right, forward (down), backward (up), diagonal</td>
<td>None</td>
</tr>
<tr>
<td>Let's Go Nuts</td>
<td>2D platformer; action</td>
<td>Jump, avoid danger, and collect nuts as you progress through the levels</td>
<td>Left, right</td>
<td>1 large button for jump</td>
</tr>
<tr>
<td>Horizon Chase Turbo</td>
<td>Racing; sports; arcade</td>
<td>Arcade racing games with various cars and tracks to select from</td>
<td>Left, right</td>
<td>Variable trigger for car acceleration</td>
</tr>
<tr>
<td>PAC-MAN Championship Edition DX+</td>
<td>Arcade; action</td>
<td>Move through the maze in this classic ghost chomping arcade game</td>
<td>Left, right, forward (up), backward (down)</td>
<td>1 large button to scare away the ghosts</td>
</tr>
<tr>
<td>Stardust Galaxy Warriors</td>
<td>Shooter; action</td>
<td>Soaring through space using various weapons to combat asteroids and enemy ships</td>
<td>Left, right, forward (up), backward (down), diagonal</td>
<td>Up to 4 buttons for weapon use</td>
</tr>
</tbody>
</table>

Measures

Quantitative Measures

The participants’ responses to a series of survey measures were collected during the session. Descriptive information was collected regarding enjoyment and current level of physical activity and video gaming. Before beginning game play, participants were shown a short video demonstrating a player using the system under 3 different conditions: sitting in a 4-legged chair, sitting in a wheelchair, and standing.

Given that higher self-efficacy is correlated with higher exercise participation [20], positive physical activity behaviors [21], AVG enjoyment [22], exercise adherence [23], exercise duration [24], and AVG approval [25], participants were asked to rate their level of task self-efficacy after watching the video. This scale (Multimedia Appendix 1) was developed based on the recommendations of Bandura et al [26] and consisted of 6 items that best represent the dimensions of the task. These dimensions equally represent the steps to the Videogame Interaction Model proposed by Yuan et al [27], which include receiving stimuli, determining response, and providing input. Each item on the scale was scored on an 11-point Likert scale from no certainty (0) to absolute certainty (10) in performing the 6 dimensions related to game play (maintaining focus, seeing and hearing game information, reacting fast, determining movement strategies, coordinating movements, and moving well). The gradations of scoring for the scale were consistent with expert recommendations for measuring self-efficacy [26]. Participants rated their self-efficacy again at the end after playing all 5 games. An average score for the 6 items was calculated, with a higher score indicating greater self-efficacy in their ability to execute game play actions.
After playing each game, participants were asked to rate their level of perceived exertion (rating of perceived exertion [RPE]) and enjoyment. For RPE, the 0 (extremely easy) to 10 (extremely hard) point adult OMNI scale was used [28]. For enjoyment, participants were presented with a visual analog scale, with anchors from “Not at All Enjoyable” on the left to “Extremely Enjoyable” on the right, and were asked to mark a spot on the line to represent their enjoyment.

After finishing all game play, participants were asked to complete 2 usability surveys. The participants’ perceived usability of the adapted video gaming controller was assessed using the System Usability Scale (SUS) and Health-IT Usability Evaluation Scale (HITUES). These 2 scales were selected to obtain general system usability data (SUS) in addition to usability data specific to AVG play using the GAIMplank (HITUES).

For the SUS, respondents answered 10 questions using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The SUS is widely used as a reliable and valid measure of a system’s usability [29-31]. Reliable results can be obtained with small samples [32], and the tool is applicable to a wide range of systems [29,33]. Using the recommended scoring guidelines, the SUS produces a score ranging from 0 to 100 [31,34]. An average SUS score of 68 corresponds to a percentile ranking of 50%, so a score >68 would be considered above average, and a score <68 is considered below average [30,33]. Adjective ratings, to help with interpretation, have been found to correlate well with the SUS scores [35].

The HITUES consisted of 20 items rated on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The HITUES has 4 sections with questions to address impact, perceived usefulness, perceived ease of use, and user control. Furthermore, the HITUES allows questions to be tailored to the system under study. For the purposes of this study, each question was phrased to assess the use of the “adapted gaming board” for “active video gaming” (Multimedia Appendix 2). Studies have demonstrated the reliability and validity of the HITUES [36-38]. One recent study suggested a cutoff score of 4.32 to indicate acceptable usability, but further validation studies are required to confirm this [39].

Qualitative Measures

Completion of the usability surveys was then followed by collection of feedback data via surveys and semistructured interviews. The questions for both were written with input from members of our engineering team to obtain feedback regarding current usability and to identify issues that needed to be addressed in future iterations and to help in the next phase of the GAIMplank development. Initially, the participants provided written responses to a series of multiple choice, yes or no, and short answer questions to obtain feedback regarding the system (Multimedia Appendix 3). Participants completed these questions on their own. After the first 8 participants, the research staff began conducting semistructured interviews with participants to facilitate more in-depth responses. Using an interview guide (Multimedia Appendix 4), the feedback questions were asked by a member of the research team in a semistructured interview format, with simple prompts (eg, please tell me more) given when appropriate. The interviews were audio recorded and later transcribed.

Results

Overview

A total of 21 adults (n=15, 71% men and n=6, 29% women) completed the usability testing. The participant characteristics are provided in Table 2. The mean age of the participants was 48.8 (SD 13.8; range 21-73) years, with 57% (12/21) White, 38% (8/21) Black, and 5% (1/21) Asian. All participants had a mobility impairment resulting from a physical disability, including stroke (9/21, 43%), spinal cord injury (3/21, 14%), amputation (3/21, 14%), cerebral palsy (2/21, 9%), spina bifida (2/21, 9%), or other (2/21, 9%). The primary modes of mobility included walking without an assistive device (7/21, 33%), cane (6/21, 29%), prosthetic leg (1/21, 5%), rollator walker (1/21, 5%), and manual wheelchair (6/21, 29%). The participants played in 1 of the following 3 ways: standing (8/21, 38%), seated in a 4-legged chair (7/21, 33%), or seated in their own manual wheelchair (6/21, 29%).

Participants reported enjoyment of leisure time physical activity, were physically active, and varied in their level of video game play (Table 3). All but 2 participants were aware of AVG. The 2 most preferred devices for playing video games included video game consoles and cell phones.

The usability data for each individual participant are reported in Table 2, with the summary group scores in Table 4. Subscale scores (usability and learning) for the SUS usability tool are also reported. The overall SUS score for all participants was 74.8 (SD 14.5), suggesting above average usability for the GAIMplank. On the basis of the adjective ratings for the SUS scale as provided by Bangor et al [35], the participant scores suggested good usability. When broken down by play style group, the average SUS scores indicated that the usability was best for those who played while seated in a chair, followed by those who played standing, and lowest for individuals who played while seated in their own wheelchairs. A similar pattern of scores was seen with the HITUES scale, with an average score of 4.3 (SD 0.3) for all participants. The HITUES scores can range from 1 to 5, with a higher score indicating greater usability.

RPE and enjoyment scores are reported for each game by play style group in Table 5. RPE scores ranged from light to moderate intensity, with the highest scores for those who played while seated in a chair. Participants rated their experience with playing each game from above average to very enjoyable.

The task self-efficacy questions were used to gauge the participants’ self-efficacy in performing certain video game play tasks using the GAIMplank system. Above average scores were reported by all 3 groups for each task (Table 6). It appears that one exposure to video game play with the GAIMplank did not diminish their task self-efficacy, and the participants felt confident in their ability to complete game video play tasks using the GAIMplank system.
Table 2. Participant characteristics and usability scores presented by play style grouping (wheelchair, seated, and standing).

<table>
<thead>
<tr>
<th>ID</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Condition</th>
<th>Primary, secondary assistive device</th>
<th>GAIMplank play style</th>
<th>Enjoy video game play (1-5)</th>
<th>Video game play (minutes per week)</th>
<th>SUS\textsuperscript{a} total score</th>
<th>HITUES\textsuperscript{b} score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>38</td>
<td>Male</td>
<td>Spinal cord injury</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>5</td>
<td>320</td>
<td>52.5</td>
<td>4.35</td>
</tr>
<tr>
<td>12</td>
<td>52</td>
<td>Male</td>
<td>Spinal cord injury</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>4</td>
<td>480</td>
<td>70</td>
<td>4.3</td>
</tr>
<tr>
<td>13</td>
<td>52</td>
<td>Male</td>
<td>Double above-knee and double below-elbow amputation</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>4</td>
<td>210</td>
<td>87.5</td>
<td>4.18</td>
</tr>
<tr>
<td>14</td>
<td>39</td>
<td>Male</td>
<td>Spina bifida</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>5</td>
<td>120</td>
<td>40</td>
<td>4.15</td>
</tr>
<tr>
<td>20</td>
<td>48</td>
<td>Female</td>
<td>Spina bifida</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>3</td>
<td>0</td>
<td>87.5</td>
<td>4.1</td>
</tr>
<tr>
<td>21</td>
<td>58</td>
<td>Male</td>
<td>Spinal cord injury</td>
<td>Wheelchair</td>
<td>Wheelchair</td>
<td>5</td>
<td>2520</td>
<td>62.5</td>
<td>3.6</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>Male</td>
<td>Stroke</td>
<td>Cane</td>
<td>Seated</td>
<td>1</td>
<td>0</td>
<td>82.5</td>
<td>4.75</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>Male</td>
<td>Stroke</td>
<td>Cane</td>
<td>Seated</td>
<td>4</td>
<td>30</td>
<td>90</td>
<td>4.5</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>Male</td>
<td>Stroke</td>
<td>Cane, walker, wheelchair</td>
<td>Seated</td>
<td>5</td>
<td>30</td>
<td>87.5</td>
<td>4.6</td>
</tr>
<tr>
<td>7</td>
<td>63</td>
<td>Male</td>
<td>Stroke</td>
<td>Cane, wheelchair</td>
<td>Seated</td>
<td>4</td>
<td>0</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>41</td>
<td>Female</td>
<td>Stroke</td>
<td>Cane, wheelchair</td>
<td>Seated</td>
<td>2</td>
<td>0</td>
<td>67.5</td>
<td>4.3</td>
</tr>
<tr>
<td>10</td>
<td>48</td>
<td>Male</td>
<td>Stroke</td>
<td>None, cane</td>
<td>Seated</td>
<td>5</td>
<td>0</td>
<td>90</td>
<td>4.58</td>
</tr>
<tr>
<td>15</td>
<td>63</td>
<td>Male</td>
<td>Spinal cord injury</td>
<td>Rollator</td>
<td>Seated</td>
<td>4</td>
<td>70</td>
<td>80</td>
<td>4.58</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>Male</td>
<td>Stroke</td>
<td>Cane, wheelchair</td>
<td>Stand</td>
<td>5</td>
<td>60</td>
<td>62.5</td>
<td>4.8</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>Male</td>
<td>Single above-knee amputation</td>
<td>Prosthetic leg</td>
<td>Stand</td>
<td>5</td>
<td>1380</td>
<td>72.5</td>
<td>4.15</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
<td>Female</td>
<td>Stroke</td>
<td>Walk</td>
<td>Stand</td>
<td>4</td>
<td>900</td>
<td>75</td>
<td>4.05</td>
</tr>
<tr>
<td>11</td>
<td>42</td>
<td>Female</td>
<td>Stroke</td>
<td>None, cane</td>
<td>Stand</td>
<td>5</td>
<td>630</td>
<td>97.5</td>
<td>4.45</td>
</tr>
<tr>
<td>16</td>
<td>29</td>
<td>Male</td>
<td>Cerebral palsy</td>
<td>None</td>
<td>Stand</td>
<td>4</td>
<td>320</td>
<td>65</td>
<td>4.15</td>
</tr>
<tr>
<td>17</td>
<td>32</td>
<td>Female</td>
<td>Single below-knee amputation</td>
<td>Prosthetic leg, wheelchair</td>
<td>Stand</td>
<td>5</td>
<td>1275</td>
<td>77.5</td>
<td>4.6</td>
</tr>
<tr>
<td>18</td>
<td>21</td>
<td>Female</td>
<td>Hydrocephalus, balance issues</td>
<td>None</td>
<td>Stand</td>
<td>4</td>
<td>10</td>
<td>85</td>
<td>4.43</td>
</tr>
<tr>
<td>19</td>
<td>40</td>
<td>Male</td>
<td>Cerebral palsy</td>
<td>None</td>
<td>Stand</td>
<td>3</td>
<td>0</td>
<td>62.5</td>
<td>4.15</td>
</tr>
</tbody>
</table>

\textsuperscript{a}SUS: System Usability Scale.

\textsuperscript{b}HITUES: Health-IT Usability Evaluation Scale.
Table 3. Participants’ enjoyment of physical activity, weekly minutes of physical activity, enjoyment of playing video games, and weekly minutes of video game play.

<table>
<thead>
<tr>
<th></th>
<th>All (n=21), mean (SD)</th>
<th>Wheelchair (n=6), mean (SD)</th>
<th>Seated in chair (n=7), mean (SD)</th>
<th>Standing (n=8), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I enjoy engaging in physical activity during my leisure time (1=strongly disagree, 5=strongly agree)</td>
<td>4.6 (0.6)</td>
<td>4.5 (0.5)</td>
<td>4.7 (0.5)</td>
<td>4.5 (0.8)</td>
</tr>
<tr>
<td>Weekday leisure-time physical activity that increases heart rate (minutes)</td>
<td>259 (177)</td>
<td>263 (230)</td>
<td>248 (171)</td>
<td>263 (163)</td>
</tr>
<tr>
<td>Weekend leisure-time physical activity that increases heart rate (minutes)</td>
<td>116 (103)</td>
<td>100 (131)</td>
<td>108 (70)</td>
<td>134 (111)</td>
</tr>
<tr>
<td>I enjoy playing video games (1=strongly disagree, 5=strongly agree)</td>
<td>4.1 (1.1)</td>
<td>4.3 (0.8)</td>
<td>3.6 (1.5)</td>
<td>4.4 (0.7)</td>
</tr>
<tr>
<td>Weekday video gaming (minutes)</td>
<td>264 (462)</td>
<td>462 (756)</td>
<td>14 (27)</td>
<td>359 (389)</td>
</tr>
<tr>
<td>Weekend video gaming (minutes)</td>
<td>138 (219)</td>
<td>204 (296)</td>
<td>4 (11)</td>
<td>213 (230)</td>
</tr>
</tbody>
</table>

Table 4. Self-reported GAIMplank usability scores.

<table>
<thead>
<tr>
<th>Game play style</th>
<th>SUS(^a) overall, mean (SD)</th>
<th>SUS: usability subscale, mean (SD)</th>
<th>SUS: learning subscale, mean (SD)</th>
<th>HITUES(^b), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All players</td>
<td>74.8 (14.5)</td>
<td>75.2 (15.5)</td>
<td>73.1 (24.8)</td>
<td>4.3 (0.3)</td>
</tr>
<tr>
<td>Wheelchair (n=6)</td>
<td>66.7 (19.0)</td>
<td>66.7 (19.2)</td>
<td>66.7 (30.3)</td>
<td>4.1 (0.3)</td>
</tr>
<tr>
<td>Seated in chair (n=7)</td>
<td>82.9 (8.6)</td>
<td>85.4 (10.8)</td>
<td>72.9 (16.6)</td>
<td>4.6 (0.1)</td>
</tr>
<tr>
<td>Standing (n=8)</td>
<td>74.7 (12.1)</td>
<td>73.8 (12.3)</td>
<td>78.1 (27.3)</td>
<td>4.3 (0.3)</td>
</tr>
</tbody>
</table>

\(^a\)SUS: System Usability Scale.
\(^b\)HITUES: Health-IT Usability Evaluation Scale.
### Table 5. Rating of perceived exertion and enjoyment rating for each game by player group.

<table>
<thead>
<tr>
<th>Game</th>
<th>Rating of perceived exertion (0-10), mean (SD)</th>
<th>Enjoyment (0-100), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All players, across games</td>
<td>4.2 (1.7)</td>
<td>72.3 (16.5)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>4.5 (1.1)</td>
<td>68.9 (12.4)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>5.1 (2.2)</td>
<td>79.5 (21.1)</td>
</tr>
<tr>
<td>Standing</td>
<td>3.4 (1.5)</td>
<td>69.4 (15.7)</td>
</tr>
<tr>
<td>Feather</td>
<td>3.4 (1.9)</td>
<td>62.7 (32.7)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>3.7 (1.4)</td>
<td>68.5 (23.9)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>4.0 (2.2)</td>
<td>69.0 (38.9)</td>
</tr>
<tr>
<td>Standing</td>
<td>2.6 (1.9)</td>
<td>52.9 (34.3)</td>
</tr>
<tr>
<td>Let’s Go Nuts</td>
<td>3.6 (2.0)</td>
<td>76.3 (17.8)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>4.0 (1.1)</td>
<td>67.3 (15.8)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>4.3 (2.9)</td>
<td>78.2 (17.7)</td>
</tr>
<tr>
<td>Standing</td>
<td>2.7 (1.5)</td>
<td>82.3 (18.9)</td>
</tr>
<tr>
<td>Horizon Chase Turbo</td>
<td>4.0 (2.0)</td>
<td>70.1 (28.5)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>4.0 (1.5)</td>
<td>65.8 (18.0)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>5.0 (2.2)</td>
<td>71.3 (45.0)</td>
</tr>
<tr>
<td>Standing</td>
<td>3.0 (1.9)</td>
<td>72.3 (18.0)</td>
</tr>
<tr>
<td>PAC-MAN</td>
<td>4.9 (2.2)</td>
<td>66.3 (24.6)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>5.3 (2.3)</td>
<td>68.8 (18.2)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>5.5 (2.7)</td>
<td>66.7 (36.8)</td>
</tr>
<tr>
<td>Standing</td>
<td>4.1 (1.8)</td>
<td>64.1 (20.6)</td>
</tr>
<tr>
<td>Stardust galaxy warriors</td>
<td>5.2 (1.9)</td>
<td>80.9 (15.2)</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>5.5 (1.4)</td>
<td>74.0 (4.2)</td>
</tr>
<tr>
<td>Seated in chair</td>
<td>6.0 (2.5)</td>
<td>91.7 (11.7)</td>
</tr>
<tr>
<td>Standing</td>
<td>4.3 (1.4)</td>
<td>77.9 (19.0)</td>
</tr>
</tbody>
</table>

### Table 6. Self-reported task self-efficacy before and after the game play session.

<table>
<thead>
<tr>
<th>Task self-efficacy question</th>
<th>All (n=21), mean (SD)</th>
<th>Wheelchair (n=6), mean (SD)</th>
<th>Seated in chair (n=7), mean (SD)</th>
<th>Standing (n=8), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintaining focus throughout a 15-minute session</td>
<td>8.8 (1.7)</td>
<td>9.6 (0.9)</td>
<td>9.5 (1.2)</td>
<td>10.0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>7.1 (1.9)</td>
<td>8.7 (1.2)</td>
<td>9.8 (0.7)</td>
<td>10.0 (0.0)</td>
</tr>
<tr>
<td>Seeing and hearing all the game information</td>
<td>9.4 (1.3)</td>
<td>9.6 (0.9)</td>
<td>9.5 (1.2)</td>
<td>9.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>9.0 (1.9)</td>
<td>9.3 (0.8)</td>
<td>9.8 (0.7)</td>
<td>9.6 (1.1)</td>
</tr>
<tr>
<td>Reacting fast enough to choose a next action</td>
<td>7.0 (2.0)</td>
<td>7.7 (1.8)</td>
<td>6.5 (2.3)</td>
<td>7.3 (1.4)</td>
</tr>
<tr>
<td></td>
<td>6.1 (1.6)</td>
<td>6.7 (1.9)</td>
<td>8.1 (1.6)</td>
<td>8.6 (1.6)</td>
</tr>
<tr>
<td>Determining strategies to move during play</td>
<td>7.0 (2.6)</td>
<td>7.7 (1.8)</td>
<td>5.8 (3.3)</td>
<td>7.0 (1.7)</td>
</tr>
<tr>
<td></td>
<td>6.0 (2.2)</td>
<td>6.7 (1.9)</td>
<td>8.6 (1.7)</td>
<td>8.9 (1.4)</td>
</tr>
<tr>
<td>Coordinating body movements to carry out a strategy</td>
<td>6.9 (2.5)</td>
<td>7.7 (1.8)</td>
<td>6.2 (3.5)</td>
<td>7.7 (1.0)</td>
</tr>
<tr>
<td></td>
<td>6.0 (1.7)</td>
<td>7.7 (1.2)</td>
<td>8.1 (1.8)</td>
<td>7.8 (2.7)</td>
</tr>
<tr>
<td>Moving well enough to maintain successful play</td>
<td>7.0 (2.4)</td>
<td>8.1 (2.0)</td>
<td>5.7 (2.4)</td>
<td>7.7 (2.1)</td>
</tr>
<tr>
<td></td>
<td>6.6 (2.4)</td>
<td>8.3 (1.2)</td>
<td>8.3 (2.0)</td>
<td>8.3 (2.5)</td>
</tr>
<tr>
<td>Overall</td>
<td>7.7 (1.7)</td>
<td>8.4 (1.1)</td>
<td>7.2 (1.8)</td>
<td>8.2 (0.8)</td>
</tr>
<tr>
<td></td>
<td>6.9 (1.9)</td>
<td>7.9 (0.8)</td>
<td>8.8 (1.1)</td>
<td>8.9 (1.4)</td>
</tr>
</tbody>
</table>
Participant Feedback

Overview

Feedback data were collected in a written response to multiple choice, yes or no, and open-ended survey questions (6/7, 86% seated and 1/7, 14% standing) or orally via a semistructured interview (5/11, 45% wheelchair; 1/11, 9% seated; and 5/11, 45% standing) conducted by the researchers. Data were missing for 1 survey participant and 2 interview participants.

From the survey feedback, when asked about the ease of mounting and dismounting the gaming board as “relatively simple,” “needs improvement,” or “difficult,” all participants rated both actions as “relatively simple.” When asked whether the gaming board was sturdy, 6 of the 7 participants reported “yes.” All participants indicated that “yes” they were able to determine where best to position themselves on the gaming board and 86% (6/7) reported that “yes” visual cues should be included to locate the central position (on the gaming board). In addition, 86% (6/7) of participants indicated that “yes,” moving their trunk (leaning) provided a responsive input for game control. Additional input functions that the participants would like to see incorporated into the gaming board included visual markers for foot placement, more games, and moving the side handrails closer. When asked to describe their overall experience on the gaming board, comments included “Positive, thank you!” “Very good,” “Something new to keep alert and enjoy,” “Entertaining,” “A lot of fun,” “It was cool to watch my character turn as I leaned left or right,” and “It was great. It was challenging because I am not a gamer, but it was a good challenge.”

Audio recordings of the interviews were transcribed verbatim. The transcripts were then reviewed independently by 2 members of the research team. From the independent reviews, themes were drawn, followed by discussion and review, with a final consensus reached on 5 main themes. These themes included accessibility of the GAIMplank system. Specific aspects highlighted by the participants were accessibility of the GAIMplank for persons with mobility impairments, usability of the GAIMplank for persons with mobility impairments, overall experience using the GAIMplank, physical activity associated with game play, and suggestions for future iterations.

Accessibility of the GAIMplank

This category examined the perceptions of the participants regarding the accessibility of the GAIMplank controller for video game play. Three specific features were discussed, including mounting and dismounting of the GAIMplank, handrails, and accessories. Most participant comments indicated adequate accessibility for video game play. None of the participants reported difficulty in mounting or dismounting the GAIMplank. The handrails provided support when needed, an overall sense of safety, and facilitated greater movement during game play by reducing the fear of falling.

Overall, the use of an assistive device during the session (ie, manual wheelchair or prosthetic limb) did not inhibit gaming activities. For participants who played seated in their manual wheelchair, blocks were placed behind their back wheels to keep them stationary for a better response to their movements. However, both participants who wore a prosthetic leg (one above the knee and one below the knee amputee) reported that they had to adjust their stance on the GAIMplank to obtain a better response.

The positioning of the external buttons and variable trigger used for game play actions such as jumping, shooting, and car acceleration was acceptable for all participants. Participants reported that the ability to adjust the height of the buttons and position them relative to their arm reach and preferred side (right, left) facilitated game play. The participants felt that the size of the buttons was suitable and that they were easy to access. One participant, who had difficulty keeping his hand on the buttons because of partial paralysis, reported that the surface friction of the buttons could be improved.

Usability of the GAIMplank

This category describes the participants’ perceptions regarding the use of the GAIMplank controller for playing video games. Specific aspects that were discussed included learning how to move for game control and responsiveness of the GAIMplank. Participants across the 3 groups acknowledged that it took some time to learn how to control and shift their weight to maneuver their character in the game. Sometimes it took time for them to feel secure to move.

Some participants reported inconsistency issues with the GAIMplank not calibrating correctly and not being responsive to their weight shifts. In some instances, recalibrating the system helped. Despite technical issues with the board during some sessions, the participants felt positive and eager to continue playing. Regardless of responsiveness issues with the GAIMplank, several participants indicated that this was the first time they were able to play such games because of accessibility issues with other AVG systems.

Overall Experience Using the GAIMplank for Video Game Play

Participants described their overall experience using the GAIMplank for video game play. Participants mainly shared positive feedback regarding their experiences with the device. As noted above, the most frequently reported barrier affecting the overall experience was the inconsistency in responsiveness of the GAIMplank system. Specific aspects highlighted by the participants regarding the GAIMplank were that the device is innovative, playing video games using the GAIMplank increases the physical activity demands of videogaming, and using the device is enjoyable. The participants expressed an interest in participating in future AVG research projects and having a device like the GAIMplank in their home setting.

“Active” Video Gaming

Several participants noted that playing video games using the GAIMplank system provided a level of physical activity. Participants described how they enjoyed the experience of moving their body to play the games instead of sitting on the couch and simply using a handheld controller or their phone. Although some frustration was experienced when the controller did not respond as expected to their movements, the participants acknowledged the value of this accessible controller in allowing them to engage in AVG play, which most of the participants...
were unable to do with current OTS products. Description of their game play experience included statements such as the following:

- I like the board because I get to move with my body versus just sitting down. Gets me up instead of just using phone. This gaming board encourages body movement. [AVGu011, walks without an assistive device, played standing]

- It was fun. It was different. With me I’m reserved to buy some of the more active games, not knowing if I can do them. To come to a controlled environment and be able to do the movements, able to sense the movements with a prosthetic. [AVGu017, walks with a prosthetic leg, played standing]

- Use the whole body. It would be great to know how many calories burned. [AVGu013, uses a manual wheelchair for mobility, played in a wheelchair]

- But at the same time its accomplishing different goals, its being more core strength and stuff than with a (hand) controller and so... (shoulder shrug). [AVGu014, uses a manual wheelchair for mobility, played in a wheelchair]

**Future Iterations**

The participants were asked to provide suggestions for improving the GAIMplank and any additional features they would like to see incorporated. The input from standing players included visual markers for foot placement, inward adjustment of handrails, handrail hook–like grip to accommodate persons without dexterity, and elevated placement of the LED matrix for better visibility. For individuals who were seated in their wheelchair, a suggestion was made to eliminate the use of the blocks behind the back wheels by adding some sort of ridge or divot to keep them from rolling.

**Researcher Observations**

During data collection for several participants (approximately 50%), technical difficulties with the GAIMplank controller occurred. The most frequent were calibration issues. During these sessions, the calibration procedure did not work as intended resulting in an inability of the device to find center, and causing difficulty for players to produce the necessary actions for game play. Recalibration and adjustments to the sensitivity and bias were attempted, but the problem was not corrected. Adjustments to the participant’s positioning on the board were also attempted, but these were not successful. Sometimes it was just one particular game during a session that would not respond to the participant’s movement. A pattern to these difficulties was not detected, and usually during the next testing session the controller worked appropriately. Specific issues based on play style (wheelchair, seated in a 4-legged chair, or standing) were not evident, except for participants who played standing and were not able to distribute their weight evenly. It appeared that the calibration procedure may not have accounted for this shift in body weight. Owing to time constraints, sufficient time was not always available to spend on troubleshooting during participant visits. Although enjoyment and RPE scores were likely affected by technical issues, participants still reported having fun and felt confident in their ability to use the system. Participants also recognized the system as providing an opportunity, typically not available to them, to engage in AVG.

**Discussion**

**Principal Findings**

Physical activity options are limited for people with mobility impairments. AVG, also known as exergaming, has the potential to provide a fun and engaging way to increase daily physical activity or reduce sedentary time. This project aimed to develop an AVG controller that was accessible and usable by individuals with mobility impairments. A sample of 21 adults with various mobility impairments was able to successfully access and play a series of standard PC video games using trunk and body movements to produce game play action. Participant usability scores (SUS and HITUES) and qualitative feedback indicated above average usability for the GAIMplank system. A previous study adapted the Wii Fit balance board controller and demonstrated successful use among people with mobility impairments with an average SUS score of 71.7 (SD 18.03) [17]. Physiological testing during game play on the adapted board indicated light-to-moderate intensity exercise for some participants [18]. This previously adapted gaming board controller was limited to use with only Wii Fit games, whereas the new GAIMplank system includes several new options for playing AVGs among people with mobility impairments and allows use with all PC games. The GAIMplank system provides a means for making games that are typically sedentary, using only the hands and fingers, more active by incorporating larger body movements. Furthermore, participants reported game play enjoyment and for some their first opportunity to engage in AVG owing to accessibility issues with OTS systems.

Looking at the average SUS scores for each participant, the lowest score was 40, well below what is considered acceptable usability. Similarly, the participant’s HITUES score (4.18) suggested below acceptable usability. The participant was a regular video game player (120 minutes per week), played seated in his wheelchair, reported moderate level enjoyment across games, game play was rated as a light-intensity exercise, and a large drop in self-efficacy was reported at the end of the session for the questions that pertained to reacting fast enough to choose the next action and moving well enough to maintain successful play. His postplay feedback highlighted issues with responsiveness, which he found frustrating. The next lowest SUS score (52.5) was also reported by a wheelchair user, which can be attributed to the GAIMplank malfunctioning, making the calibration off and the games moving slow. Regardless, this participant who was physically active (960 minutes per week) and a frequent video game player (320 minutes per week) reported above average enjoyment scores, game play as light-intensity exercise, with an improvement in overall self-efficacy after using the GAIMplank system. Of the 4 other participants who reported less than optimal usability (62.5-65), 3 (75%) were standing players who were all physically active, with varying levels of weekly video game play. Two players also had low HITUES scores (4.15) and reported moderate...
enjoyment, light-intensity exercise, and above average self-efficacy. In both cases, technical issues were encountered with the GAIMplank system during the testing session, resulting in poor responsiveness and making the games very difficult to play. The other standing participant typically used a rollator to assist with mobility but was able to play without it. His average enjoyment score was high (86) with light-intensity RPE ratings and maximum self-efficacy scores. Although his SUS score was low (62.5), his HITUES score (4.8) was the highest among all participants. These scores may suggest that although the participant perceived general usability (SUS) issues with the system, he recognized the usability of the GAIMplank as an adapted controller to provide an option for active video game play for individuals with mobility impairments.

At the other end of the spectrum, 7 players reported “excellent” usability (SUS ≥85) for the GAIMplank system. Of those, 2 played in their wheelchair, 2 played standing, and 3 played while seated in the chair. For the standing and seated players, the HITUES scores reflected acceptable usability. All 3 seated players (poststroke) spent little to no time playing video games each week but were physically active, with 2 reporting >400 minutes per week of physical activity. The average enjoyment scores were high, the RPE scores varied, and all were reported to have certainty in their ability to conduct various video game play tasks using the GAIMplank system. Both standing participants were physically active (120 minutes per week); one spent a lot of time playing video games each week (630 minutes), whereas the other played very little, if any. Average enjoyment scores were higher for the person who typically spent a lot of time playing video games each week, whereas RPE was higher for the other participant. Both participants reported high self-efficacy in using the GAIMplank system for video game play. Both participants who played seated in their wheelchairs with high SUS scores were physically active (120 minutes per week); one spent a lot of time playing video games each week (210 minutes), whereas the other one played none. The HITUES scores were lower than those of other participants who reported high SUS scores. Both players enjoyed the games, with RPE scores indicating light-to-moderate intensity exercise. The self-efficacy scores were high for both players at the end of the session.

The results demonstrate the usability of a newly developed adapted gaming board for AVG play, but attention is needed to improve the stability of responsiveness and to address issues experienced by different subgroups of individuals with mobility impairments. Considerations for the engineering team during the next phase of development include different weight distribution patterns during game play by individuals wearing a prosthetic leg, varying movements associated with spasticity among certain players, foot placement outside the board (ie, on stool in front) for some seated players, and overall improvement to better accommodate wheelchair play. Overall, the ability of the adapted controller to provide a fun and accessible option for AVG play was clearly recognized by the players, suggesting the need for additional development and research in this area.

Study Limitations

The heterogeneous nature of the sample makes it difficult to determine whether different aspects of the GAIMplank better suited one group over another. The number of participants within each play style group was small and consisted of individuals with various mobility impairments. In addition, participants were recruited from a community-based health and fitness center, so all had experience with physical activity and were currently active with reported weekly physical activity minutes ranging from 60 to 960 (mean 375, SD 257) minutes. Therefore, the results cannot be generalized to more sedentary populations. Previous gaming experience was captured, but its influence on system usability or their game play experience could not be teased out. With regard to the specific games used for the sessions, game selection was limited and may not have appealed to some participants. In addition, sufficient time to learn how best to move for each game was not provided. Initially, qualitative feedback data were not collected from participants. As the need for more in-depth user input was recognized, semistructured interviews were conducted to collect richer feedback data. This change occurred after most of the seated players had completed testing; therefore, limited qualitative data were available from this group.

Future Considerations

Continued testing and refinement of the GAIMplank system are underway. The usability testing results were discussed with the engineering and design teams, and a new wave of refinements is planned with a focus on improving the stability of the platform. Specifically, elimination of the Raspberry Pi and addition of a custom-fabricated PCB board are being considered. Following further refinement of the system, the plan is to conduct a feasibility study to examine the use of the GAIMplank for an AVG exercise intervention. The outcomes of interest will include various physiological and psychosocial measures. Other factors to be considered include game play selection and preferences, previous gaming experience, participants’ physical activity level, and single versus multiplayer options. In addition, consideration of factors that affect game play-by-play style (eg, seated and standing) needs to be examined. The potential of integrating other gaming controllers is also of interest.

Conclusions

The GAIMplank, an adapted gaming controller, allowed people with various mobility impairments to engage in AVG. Using the GAIMplank system, standard PC video games can be played while sitting, standing, and in a manual wheelchair. Participants reported above average usability for the GAIMplank and found game play to be enjoyable and provide light-to-moderate intensity exercise. Further design and development work to refine the device is recommended.
Acknowledgments

The content of this manuscript was developed by the Rehabilitation Engineering Research Center RecTech under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (grant 90REGE0002). National Institute on Disability, Independent Living, and Rehabilitation Research is a center within the Administration for Community Living, Department of Health and Human Services. The authors are extremely grateful for the assistance of Jake Davis throughout the preliminary testing phases and would also like to thank John Franks and Amy Belcher for their contributions. The authors would like to thank the University of Alabama at Birmingham Engineering and Innovative Development group, including Andres Morales, Emma Schmidt, Emma Latham, Alex Plazas, and Martin Holland for the design and development of the adapted gaming board. The authors are also grateful for the contributions of Sean Simmons from Objective Design.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey to assess participant pregame and postgame play self-efficacy. [DOCX File, 18 KB - games_v11i1e38484_app1.docx ]

Multimedia Appendix 2
Health-IT Usability Evaluation Scale used to collect usability data specific to active video gaming using the GAIMplank. [DOCX File, 23 KB - games_v11i1e38484_app2.docx ]

Multimedia Appendix 3
Survey questions used to gather additional feedback from participants regarding the GAIMplank. [DOCX File, 18 KB - games_v11i1e38484_app3.docx ]

Multimedia Appendix 4
Questions used to guide semistructured interview with participants to collect feedback data at the end of game play using the GAIMplank system. [DOCX File, 18 KB - games_v11i1e38484_app4.docx ]

References


15. Chu CH, Biss RK, Cooper L, Quan AM, Matulis H. Exergaming platform for older adults residing in long-term care homes: user-centered design, development, and usability study. JMIR Serious Games 2021 Mar 09;9(1):e22370 [FREE Full text] [doi: 10.2196/22370] [Medline: 33687337]


Abbreviations

AVG: active video gaming
HITUES: Health-IT Usability Evaluation Scale
OTS: off-the-shelf
PCB: printed circuit board
PoC: proof-of-concept
PoP: proof-of-product
RPE: rating of perceived exertion
SUS: System Usability Scale

Edited by N Zary; submitted 04.04.22; peer-reviewed by L Weinert; comments to author 07.05.22; revised version received 02.07.22; accepted 11.10.22; published 10.01.23.

Please cite as:
Malone LA, Mendonca CJ, Mohanraj S, Misko SR, Moore J, Brascome JM, Thirumalai M
Usability of the GAIMplank Video Game Controller for People With Mobility Impairments: Observational Study
JMIR Serious Games 2023;11:e38484
URL: https://games.jmir.org/2023/1/e38484
doi:10.2196/38484
PMID:36626195

©Laurie A Malone, Christen J Mendonca, Sangeetha Mohanraj, Samuel R Misko, Joseph Moore, James Michael Brascome, Mohanraj Thirumalai. Originally published in JMIR Serious Games (https://games.jmir.org), 10.01.2023. 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.
Impact of Gamification on Consumers’ Favorability in Cause-Related Marketing Programs: Between-Subjects Experiments

Yanhe Li\textsuperscript{1*}, PhD; Yanchen Li\textsuperscript{2*}, PhD; Xiu Zhou\textsuperscript{3*}, PhD; Kunshu Ma\textsuperscript{4*}, PhD

\textsuperscript{1}School of Management, Zhejiang Shuren University, Hangzhou, China
\textsuperscript{2}School of Business, Macau University of Science and Technology, Macau S.A.R, China
\textsuperscript{3}School of Accounting and Finance, Beijing Institute of Technology, Zhuhai, China
\textsuperscript{4}School of Management, Southwest Minzu University, Chengdu, China
\textsuperscript{*}all authors contributed equally

Corresponding Author:
Yanchen Li, PhD
School of Business
Macau University of Science and Technology
Avenida Wai Long
Taipa
Macau S.A.R, 999078
China
Phone: 86 88972037
Email: ycli@must.edu.mo

Abstract

Background: Successful cause-related marketing (CRM) campaigns can help companies stand out from their competitors; however, CRM may not have pleasant outcomes, even if it receives substantial investment.

Objective: This research aimed to investigate how gamified CRM projects influence consumers’ favorability.

Methods: We introduced 3 different CRM projects in 3 different studies. Every project had 2 versions according to the level of gamification, and participants were randomly assigned into these 2 groups. Additionally, we used a 2 (gamification: lower, higher) 2 (rules presentation: without visual cues, with visual cues) between-subjects design to test the moderation role of rules presentation in gamified CRM projects.

Results: In Study 1, we identified that the highly gamified CRM program induces more enjoyment ($F_{1,139}=21.11$, $P<.001$) and higher favorability ($F_{1,139}=14.57$, $P<.001$). Moreover, we found that enjoyment played a mediation role between gamification and favorability ($P<.001$) in Study 2. In addition, the results of Study 3 indicated rules presentation in a gamified CRM program can moderate the indirect effect of gamification on favorability via enjoyment (index of the moderated mediation: 95% CI –1.12 to –0.10; for rules presentation with visual cues: 95% CI 0.69 to 1.40; for rules presentation without visual cues: 95% CI 0.08 to 0.83).

Conclusions: Overall, this research contributes to the CRM literature and suggests gamification is an effective way of managing CRM campaigns.

(JMIR Serious Games 2023;11:e35756) doi:10.2196/35756

KEYWORDS
cause-related marketing; gamification; enjoyment; favorability; business; marketing; gamified

https://games.jmir.org/2023/1/e35756
**Introduction**

**Background**

Corporate social responsibility practices are currently popular for firms to enable themselves to stand out from their competitors and achieve sustainable performance [1]. Generally, companies undertake cause-related marketing (CRM) as one solution to show their commitment to corporate social responsibility [2]. Successful CRM campaigns can help companies elicit positive attitudes [3] and emotional bonds between them and their customers [4]. However, CRM may not have pleasant outcomes, even if it receives substantial investment [5,6]. It is therefore critical to understand the related predictors so that practitioners can develop CRM activities that will have desirable outcomes.

Gamification is a new and effective way to generate CRM success. Recently, some companies have applied gamification to their CRM practices. For example, Alipay, one of the famous e-payment applications in China, launched a gamified CRM scheme called “Ant Forest” in 2016. Users grow virtual trees in Ant Forest by collecting game points. As users complete the game tasks, they earn points for taking buses, walking daily steps, or using online payment in a low-carbon lifestyle. After obtaining enough points, users can turn their virtual trees into real ones and thus contribute to environmental protection. The hedonic benefit provided by a gamified system motivates users to take action. However, to our best knowledge, few studies discuss gamified CRM and its impact on people's attitudes or behaviors [7]. In this article, we advanced previous research and performed empirical studies to answer related questions.

**Enjoyment and Favorability**

Consumers are not totally rational. As they make purchase decisions, they involve their fantasies, emotions, or other experiential perspectives rather than just processing the information [8].

Gamification is thus the popular choice for practitioners to satisfy consumers with utilitarian values, hedonic values, and social values [9]. On one hand, the concept of gamification originates from “game.” Gamification indicates the use of game design elements, game mechanisms, or games in nongame contexts [10]. It aims to provide consumers with positive emotional and involved experiences (ie, gameful experiences) [11]. On the other hand, some research (eg, Jung et al [12], Kim and Ahn [13], and Sailer et al [14]) has proved that game design elements (eg, points, badges, and leaderboards) can not only evoke users’ intrinsic motivations and yield performance gains but also internalize extrinsic motivations (eg, external rewards).

Therefore, we believed that, compared with traditional CRM activities, consumers enjoy more hedonic values and feel as if they are more involved in gamified activities, leading to the following hypothesis: (H1) Gamification has a positive influence on enjoyment.

Consumers do not adopt CRM activities casually [6]. Consumers’ personal evaluation of or beliefs around the cause can influence the success of CRM [15]. Thus, in this study, we defined favorability as consumers’ positive attitudes toward CRM activities [16,17], to measure the success of CRM programs.

Generally, CRM campaigns work when consumers benefit from feeling pleasure during participation [18]. We posited that consumers are more willing to accept gamified CRM activities, since they can offer more opportunities for consumers to interact with the cause and the perceived enjoyment from gamified CRM campaigns positively influence consumers’ favorability.

Extant research suggests that positive emotions can arouse individuals’ positive attitudes [19]. Especially in the context of CRM, hedonic products that are associated with positive emotional processing are more likely to provoke emotional contagion; individuals thus have more favorable attitudes toward the CRM programs [20].

Although studies examining the relationship between enjoyment and favorability in the context of gamified CRM are lacking, various studies (eg, Yang et al [21], Catalán et al [22], and Hwang and Choi [23]) validated the positive influence of perceived enjoyment from gamification on positive attitudes, such as consumers’ brand attitudes [24], continuous use intention [25], and brand loyalty [26].

Therefore, we hypothesized that (H2) enjoyment has a positive influence on favorability and mediates the relationship between gamification and favorability.

**Rules Presentation**

Gamification embraces the nature of games. Importantly, rules are the core factors of games [27]. Containing various game mechanics [28], rules stipulate how to achieve the goal of a game and the outcome of each trial [29].

In this article, we defined rules presentation as the way of presenting the rules of a CRM project, with or without visual cues. We believe that the way rules are presented can influence people’s enjoyment. Rules with visual cues tend to help people to process information fluently and thus induce enjoyment at a higher level. According to processing fluency theory, the ways of processing information vary from the degrees of effort and speed [30]. With a fluent process, it is faster and easier to make judgments; however, a process that is not fluent has the opposite effect.

Generally, individuals prefer a stimulus that is easier to perceive and tend to favor it [31-33]. Thus, fluency always results in positive outcomes [34]. Past research has shed light on that fact that fluency of information processing (eg, rules processing) leads to consumers’ positive affective responses (eg, Janiszewski [35] and Gambelin et al [36]) or leads to purchase intentions (eg, Zhang et al [37], Jaud and Melnyk [38], and Wang et al [39]).

Meanwhile, fluent processes are always involved in visual categorization [40,41]. When faced with experiential-attribute (ie, affective or sensory) targets, consumers are more likely to receive stimuli and respond immediately, without elaborating or reasoning [42].

Therefore, we hypothesized that, if the rules of gamified projects are performed with visual cues (eg, in figurative or other sensory ways), consumers tend to perceive higher enjoyment. On the
contrary, however, when the rules are presented in plain and
monotonous literal words, consumers perceive less or even no
enjoyment from gamified CRM programs. We thus predicted
that (H3) visual cues in rules presentation moderate the
relationship between gamification and enjoyment. People tend
to have a higher level of enjoyment under the condition with
visual cues (vs the condition without visual cues).

The conceptual model is illustrated in Figure 1.

Overall, this research aimed to test if consumers have a higher
level of favorability to respond in the context of gamification
and try to explain why gamification strategy can work.

Specifically, we hypothesized that gamification has a positive
influence on enjoyment and consumers’ favorability toward
gamified CRM programs and that enjoyment mediates the
relationship between gamification and favorability. Moreover,
we also hypothesized that visual cues in rules presentation can
moderate the relationship between gamification and enjoyment.
People tend to have a higher level of enjoyment under the
condition with visual cues (vs the condition without visual cues).
We described 3 studies in detail to test our hypotheses. Finally,
we addressed the theoretical and managerial implications of
this research and put forward some future research directions
as suggestions.

Figure 1. Conceptual model.

Methods

Study Design in Study 1
We created 2 videos to introduce the Ant Forest of Alipay. The
higher gamification version illustrated the real-life version of
Ant Forest. For example, users can gain points by finishing
certain assignments or games or stealing from friends. After
accumulating enough points, they can turn their virtual plants
into real ones. However, the lower gamification version
presented a fictitious Ant Forest interface. We deleted all the
add-in games. Specifically, users could only accumulate points
from conducting eco-friendly behaviors (eg, using public transit)
but could not accumulate points from other highly gamified
methods (eg, finishing games or interactions with friends). Since
the basic rules were retained, we regarded the fictitious version
as a lower gamified program. To minimize noise, we kept 2
versions of the same visual designs (eg, images) and presentation
length (ie, 30 seconds).

A pretest was first conducted to show that the 2 versions of Ant
Forest were not significantly different in relevant factors except
for the level of gamification. In the pretest, we told the
participants to evaluate the given Ant Forest programs.
Participants were randomly assigned to 1 of 2 conditions. We
performed a confounding check of brand trust, brand attitude,
and brand familiarity (7-point scales; see Multimedia Appendix
1 for details [43-45]). We then asked participants to rate the
perceived gamification level. The measurement was adapted
from that of Hwang and Choi [23]: “This Ant Forest program
entails a game component.” (1=Strongly Disagree; 7=Strongly
Agree).

After the pretest, we collected a new data sample to conduct
our main study. In the main study, we first provided 2 versions
of the Ant Forest introduction videos and asked participants to
rate their willingness to favor the programs. The measurement
was adapted from that of Speed and Thompson [46]: “This form
of Ant Forest makes me feel more favorable toward the
program.” “This form of Ant Forest would improve my
perception of the program.” “This form of Ant Forest would
make me like the program more.” (1=Strongly Disagree; 7=Strongly
Agree; =.78). Second, to measure enjoyment, we
used the scale from Höllig et al [47]: “I would find the program
presented in the video game enjoyable.” “I would find the
program presented in the video game enjoyable and pleasant.”
“I would find the program presented in the video game enjoyable
and exciting.” “I would find the program presented in the video
game enjoyable and interesting” (1=Strongly Disagree; 7=Strongly
Agree; =.81). Third, we asked the participants to
rate the perceived gamification level for our manipulation check.
Finally, we also performed a confounding check.

Study Design in Study 2
In Study 2, we focused on the relationships among gamification,
enjoyment, and favorability and tested if enjoyment plays the
role of mediator (H2). We proposed that individuals are more
likely to perceive higher enjoyment from a gamified CRM
program and in turn have higher favorability toward the program
(H2). In Study 2, we created a new stimulus to test the
hypothesis. Moreover, to provide generalizability for our results
related to H1, we also tested them in this study.

Similar to Study 1, we created 2 videos to introduce a CRM
program sponsored by Douyin. Douyin (also known as “TikTok”
in western countries) is a popular video-sharing social network
service in China. The program is named “Dream Fairy” and aims to donate books to children. The higher gamification version was the real-life version of Dream Fairy in which users can gain points by watching or publishing instant videos, participating in question-and-answer games, or interacting with friends. When they have accumulated enough points, they can make their donations in reality. On the contrary, the lower gamification version presented a fictitious Dream Fairy interface. We deleted all the games. Specifically, users could only accumulate points from watching instant videos or lives and from publishing instant videos. They could not accumulate points from other highly gamified methods (eg, playing games or social interactions). We also kept 2 versions of the same image designs and presentation duration (ie, 30 seconds) to minimize noise in the data. In Study 2, we also performed a pretest and a main study to verify the internal validity.

**Study Design in Study 3**

In Study 3, we conducted the mediation analysis again as a robustness check and explored the moderation role of visual cues between gamification and favorability.

We proposed that rules presentation moderates the effect of gamification on enjoyment and finally influences individuals’ favorability toward CRM (H3). In Study 3, we provided the rules to participate in a CRM either with or without visual cues. Presumably, people tend to process rules with visual cues more fluently and thus perceive more enjoyment. Therefore, we anticipated that, when rules are presented with visual cues, a higher (vs lower) gamified CRM facilitates higher favorability.

In Study 3, we chose a CRM program called “Protect Pandas” as the stimuli material. This CRM program aims to protect pandas and is sponsored by Weibo, a quite trendy mobile application in China.

Similar to previous studies, we created 2 videos to introduce the CRM program. The higher gamification version was the true version in reality. For example, users can gain points by logging into the app; publishing, reading, sharing, or liking posts; commenting; or interacting with friends. After accumulating enough points, Weibo helps users to plant actual bamboo to feed pandas. On the contrary, the lower gamification version presented a fictitious interface. The higher game elements were all deleted. Users could not accumulate points from social interactions with friends but only from daily routine behaviors in the app (eg, publishing or reading posts). We kept 2 versions in the same image designs and presentation duration (ie, 30 seconds) to avoid noise in the data. A pretest and a main study were also conducted in Study 3.

**Ethical Approval**

The university institutional review board at Zhejiang Shuren University approved all data collection procedures and scales we used (20221225).

**Data Source**

All the data in this research were collected in Credamo, which is an online survey platform. The questionnaires were randomly distributed to a registered consumer panel by the platform system, and researchers can collect data in exchange for monetary compensation. All the studies performed in this research were in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). All participants provided written consent to participate in the research and were informed that they had the ability to opt out.

## Results

**Results of Study 1**

**Pretest**

We recruited 50 participants from Credamo. The results revealed that there were no significant differences between the 2 conditions for brand trust (lower gamification mean 6.19; higher gamification mean 6.42; $P=.17$), brand attitude (lower gamification mean 6.04; higher gamification mean 6.22; $P=.38$), or brand familiarity (lower gamification mean 6.12; higher gamification mean 6.35; $P=.41$). As expected, levels of perceived gamification were significantly different between the 2 conditions (lower gamification mean 3.87; higher gamification mean 5.91; $F_{1,48}=31.63$, $P<.001$).

**Main Study**

We recruited 141 participants from Credamo; 78 (55.3%) were female, and 129 (91.5%) were aged from 21 years to 40 years. All the confounding variables (eg, brand trust, brand attitude, and brand familiarity) showed no significant differences ($P=.09$, $P=.07$, $P=.19$, respectively). Since all the participants were randomly separated into groups, we did not discuss any covariates (eg, age) further. We also performed the manipulation check on the variable of perceived gamification. The result revealed that the levels of perceived gamification were different between the 2 groups (lower gamification mean 4.89; higher gamification mean 5.83; $F_{1,139}=32.63$, $P<.001$).

We then took enjoyment as a dependent variable and conducted a similar analysis. The result showed that gamification has a positive influence on enjoyment (lower gamification mean 5.44; higher gamification mean 6.04; $F_{1,139}=21.11$, $P<.001$; partial $\eta^2=0.13$), which means the higher gamified program leads to higher enjoyment than the lower gamified program.

We finally took favorability as the dependent variable and conducted a similar analysis. The result showed that gamification has a positive influence on favorability (lower gamification mean 5.63; higher gamification mean 6.13; $F_{1,139}=14.57$, $P<.001$; partial $\eta^2=0.09$), which means the higher gamified program leads to higher favorability than the lower gamified program.

The results of Study 1 supported H1. The highly gamified CRM program induces more enjoyment and higher favorability.

**Results of Study 2**

**Pretest**

A pretest was conducted with 42 participants recruited from Credamo. Participants were randomly assigned to 1 of 2 conditions and were instructed to evaluate the given programs. We also performed a confounding check of brand trust, brand attitude, and brand familiarity. The results revealed that there...
were no significant differences between the 2 conditions for brand trust (lower gamification mean 5.66; higher gamification mean 5.66; \( P = .98 \)), brand attitude (lower gamification mean 6.11; higher gamification mean 5.91; \( P = .40 \)), or brand familiarity (lower gamification mean 6.05; higher gamification mean 6.13; \( P = .79 \)). However, perceived gamification differed between the 2 conditions (lower gamification mean 5.83; \( F_{1,40} = 10.81, P = .002 \)).

**Main Study**

We recruited 114 participants from Credamo; 55 (48.2%) were female, and all the participants were aged from 21 years to 40 years. We randomly assigned them to view 1 of the 2 videos. After viewing, we first asked the participants to rate the perceived gamification and then asked them to rate the perceived enjoyment (\( \alpha = .76 \)) and favorability (\( \alpha = .87 \)) toward the Douyin CRM program. Confounding variables (ie, brand trust, brand attitude, and brand familiarity), as in Study 1, were also checked, and they showed no significant differences between the 2 groups (\( P = .08, P = .14, P = .12 \), respectively). All the measures were the same as those in Study 1.

We performed the manipulation check on the variable of perceived gamification in Study 2, replicating the results of the pretest. The 2 videos were significantly different in levels (lower gamification mean 4.36; higher gamification mean 5.77; \( F_{1,112} = 33.87, P < .001 \)).

As in Study 1, we conducted a series of analyses of variance (ANOVA) on the dependent variables. The results showed that gamification positively influenced enjoyment (lower gamification mean 5.21; higher gamification mean 5.94; \( F_{1,112} = 19.57, P < .001 \); partial \( \eta^2 = .15 \)), with higher gamification leading to more enjoyment. The results also revealed that gamification had a positive impact on favorability (lower gamification mean 5.67; higher gamification mean 6.05; \( F_{1,112} = 11.18, P = .001 \); partial \( \eta^2 = .09 \)), with higher gamification leading to higher favorability.

To test H2, we conducted a mediation analysis using the bootstrap method (5000 bootstrap samples) in the mediation package of R. The results indicated that the effect of gamification on favorability was fully mediated by enjoyment. As Figure 2 illustrates, the total effect of gamification on favorability was significant (\( \beta = .38, P < .001 \)). Controlling for enjoyment, gamification also had a significant effect on favorability (\( \beta = .36, P < .001 \)). However, controlling for enjoyment, gamification no longer had a significant influence on favorability (\( \beta = .12, P = .17 \)). The indirect effect of gamification on favorability through enjoyment was 0.73 \( \times \) 0.36 = 0.26 (\( P < .001 \)), with the 95% CI ranging from 0.16 to 0.63, excluding 0. Thus, H2 was supported.

Study 2 not only reinforced H1 but also gave support to our hypothesis on the mediation role of enjoyment (H2) between gamification and favorability.

**Results of Study 3**

**Pretest**

We performed a pretest with 30 participants recruited from Credamo. Participants were randomly assigned to 1 of 2 conditions, and there were no significant differences between the 2 groups in brand trust (lower gamification mean 5.66; higher gamification mean 5.48; \( P = .58 \)), brand attitude (lower gamification mean 5.55, higher gamification mean 5.73; \( P = .66 \)), or brand familiarity (lower gamification mean 6.14; higher gamification mean 6.06; \( P = .84 \)). However, perceived gamification differed between the 2 conditions (lower gamification mean 5.21; higher gamification mean 6.12; \( F_{1,28} = 8.62, P = .007 \)).

**Main Study**

Study 3 used a 2 (gamification: lower, higher) \( \times \) 2 (rules presentation: without visual cues, with visual cues) between-subjects design. We recruited a new sample of 250 participants from Credamo; 128 (51.2%) were female, and all the participants were aged from 21 years to 40 years. In the videos, we first presented the rules (with or without visual cues) for “Protect Pandas” and then showed the interfaces of the program. After viewing, the participants were asked to rate the perceived gamification, enjoyment, and favorability toward the CRM program as in the previous studies. A confounding check was conducted and revealed no significant differences (\( P = .30, P = .78, P = .70 \), respectively).

---

**Figure 2.** Path coefficients of the mediation model. *****P < .001.**
Similar to the results of the pretest, the results of the manipulation check on the variable of perceived gamification showed that the 2 videos were significantly different in levels (lower gamification mean 5.16; higher gamification mean 5.72; $F_{1,248}=19.48$, $P<.001$).

We first conducted an ANOVA with enjoyment as the dependent variable and gamification and rules presentation as the independent variables. The results confirmed the main effect of gamification on enjoyment (lower gamification mean 4.71; higher gamification mean 5.56; $F_{1,246}=30.06$, $P<.001$; partial $\eta^2=0.10$) and the main effect of rules presentation (without visual cues mean 4.86; with visual cues mean 5.43; $F_{1,246}=13.91$, $P<.001$; partial $\eta^2=0.05$). These simple effects were also qualified by a significant gamification rules presentation interaction effect on enjoyment ($F_{1,246}=4.98$, $P=.03$; partial $\eta^2=0.02$). The planned contrast revealed that, when rules were presented with visual cues, higher gamification led to more enjoyment than did lower gamification (lower gamification mean 4.81; higher gamification mean 6.02; $F_{1,246}=30.92$, $P<.001$). Meanwhile, when rules were presented without visual cues, there was also a significant difference between the 2 groups (lower gamification mean 4.60; higher gamification mean 5.11; $F_{1,246}=5.43$, $P=.02$). See Figure 3.

We next performed an ANOVA with favorability as the dependent variable and gamification and rules presentation as the independent variables. The results confirmed the main effect of gamification on favorability (lower gamification mean 5.03; higher gamification mean 5.86; $F_{1,246}=29.93$, $P<.001$; partial $\eta^2=0.10$) and the main effect of rules presentation (without visual cues mean 5.21; with visual cues mean 5.72; $F_{1,246}=11.83$, $P<.001$; partial $\eta^2=0.04$). These simple effects were also qualified by a significant gamification rules presentation interaction effect on favorability ($F_{1,246}=5.08$, $P=.03$; partial $\eta^2=0.02$). The planned contrast revealed that, when rules were presented with visual cues, higher gamification led to more favorable attitudes toward the CRM program than did lower gamification (lower gamification mean 5.11; higher gamification mean 6.29; $F_{1,246}=30.92$, $P<.001$). Meanwhile, when rules were presented without visual cues, there was also a significant difference between the 2 groups (lower gamification mean 4.95; higher gamification mean 5.44; $F_{1,246}=5.29$, $P=.02$). See Figure 4.

**Figure 3.** Mean (SD) enjoyment in the higher gamification group compared with the lower gamification group.
Finally, we estimated our moderated mediation model (in the mediation package of R; 5000 bootstrap samples) to test whether the visual cues in rules representation moderate the underlying process via enjoyment. The model used gamification as the independent variable, enjoyment as the mediator, and visual cues as the moderator. Similar to Study 1 and Study 2, gamification positively influenced enjoyment ($\beta=.51$, $t_{246}=2.33$, $P=.02$). Moreover, the interaction between gamification and visual cues was significant for enjoyment ($\beta=.69$, $t_{246}=2.23$, $P<.03$). Enjoyment, in turn, facilitated favorability ($\beta=.87$, $t_{245}=30.80$, $P<.001$). Visual cues significantly moderated the indirect effect of gamification on favorability via enjoyment (index of moderated mediation: 95% CI –1.12 to –0.10; for rules presentation with visual cues: 95% CI=0.69 to 1.40; for rules presentation without visual cues: 95% CI 0.08 to 0.83). However, the direct path from gamification to favorability was not significant ($P=.61$) nor was its moderation by visual cues (95% CI –0.39 to 0.20).

In summary, the findings from Study 3 supported our hypotheses (H1 to H3). Gamified CRM projects can induce more enjoyment and higher favorability. Importantly, when rules were presented with visual cues, the impact of gamification on enjoyment was more positive than when without visual cues. The results highlight the notion that consumers perceive enjoyment from not only game design elements of the CRM program but also the rules presentation.

**Discussion**

Across 3 studies, we showed that gamified CRM programs can increase consumers’ perceptions of enjoyment and, in turn, enhance their favorability toward the program (Studies 1 and 2). We also showed that the relationship between gamification and enjoyment is moderated by the visual cues of rule presentation (Study 3). Our findings have several implications for marketing research and practice.

Theoretically, our research contributes to the CRM literature by highlighting the impact of gamification. It is not always easy to get consumers’ responses to or willingness to participate in a CRM program (eg, Chuah et al [6] and Jun et al [48]). However, in this article, we showed that gamification is an effective alternative for CRM projects to gain more favorability by improving consumers' enjoyment. Moreover, our work adds to the literature exploring the impact of visual design on consumers’ psychological mechanisms. The results of our study indicate that, when rules are presented with visual cues, consumers are more likely to perceive higher enjoyment and, in turn, feel greater favorability.
Additionally, the current research yields some implementable managerial implications. On one hand, managers can gamify their CRM programs to enhance consumers' enjoyment and favorability. For example, practitioners embed not only game design elements but also games into their projects. On the other hand, some gamified CRM programs merely present rules in monotonous words (eg, Dream Fairy in Study 2); however, according to our research, practitioners are better to present their rules with visual cues (even in a dynamic way) to induce more enjoyment.

We provide a range of evidence for our model, but there are still some limitations for future research. First, we only discussed the psychological mechanism of enjoyment between gamified CRM programs and favorability. However, there may exist multiple mediators. For example, engagement is another mediator to investigate. Since gamified CRM programs are possibly more interesting than ordinary CRM programs and provide various game design elements with which to interact, consumers may be more easily engaged [49,50]. When highly engaged, consumers are more inclined to believe the cause is connected with their lives [51] and thus are more willing to favor it [52]. Furthermore, engagement may increase the perception of one's personal role in contributing to the cause [18], which evokes an intention to participate. Second, in this article, we performed studies online and merely introduced CRM programs via videos. However, for more strict control and data collection, future studies can choose the context of a laboratory and provide opportunities for participants to conduct operations on certain programs. Finally, as the experience with gamification has not been defined clearly, the current manipulation check is limited. Future studies can develop scales for the gamification experience.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Table measurement items.
[DOCX File, 13 KB - games_v11i1e35756_app1.docx ]

References


Abbreviations

ANOVA: analysis of variance
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CRM: cause-related marketing

Edited by N Zary, G Eysenbach; submitted 16.12.21; peer-reviewed by T Baranowski, AM Bezabih; comments to author 27.05.22; revised version received 01.10.22; accepted 01.12.22; published 10.01.23.

Please cite as:
Li Y, Li Y, Zhou X, Ma K
Impact of Gamification on Consumers’ Favorability in Cause-Related Marketing Programs: Between-Subjects Experiments
JMIR Serious Games 2023;11:e35756
URL: https://games.jmir.org/2023/1/e35756
doi: 10.2196/35756
PMID: 36625196

©Yanhe Li, Yanchen Li, Xiu Zhou, Kunshu Ma. Originally published in JMIR Serious Games (https://games.jmir.org), 10.01.2023. 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.
Finding Effective Adjustment Levels for Upper Limb Exergames: Focus Group Study With Children With Physical Disabilities

Martina Eckert1*, PhD, ING; Beatriz Domingo Soria2*, PT; Noelia Terroso Gil2*, PT

1Group of Acoustics and Multimedia Applications, Department of Audiovisual Engineering, Universidad Politécnica de Madrid, Madrid, Spain
2Department of Physiotherapy, Primary School, Centro de Educación Infantil y Primaria Pinar de San José, Madrid, Spain
*all authors contributed equally

Corresponding Author:
Martina Eckert, PhD, ING
Group of Acoustics and Multimedia Applications
Department of Audiovisual Engineering
Universidad Politécnica de Madrid
Calle Alan Turing 3
Madrid, 28031
Spain
Phone: 34 91 067 87 62
Email: martina.eckert@upm.es

Abstract

Background: We developed the Blexer system consisting of a database and a web interface for therapists that can host different types of adaptive and personally configurable virtual reality exergames based on Kinect (Microsoft Corp) motion capture to provide entertaining exercises for children with motor disabilities. It allows for parameter adjustment and the monitoring of results remotely, thereby providing a useful tool to complement traditional physical therapy sessions with home exercises.

Objective: The aim of this study was to observe the motor benefits achieved through the use of a video exergame and the importance and implications of correctly setting the game’s difficulty parameters.

Methods: This was an observational case study of 6 children with different physical disabilities receiving physical therapy at school combined with the use of a fully configurable exergame under research that forms a part of the Blexer environment. The game integrates 4 repeatedly appearing upper limb exercises with individually adjustable difficulties (intermittent arm rising, arm forward and backward movement, rising and holding of one arm, and trunk control in all directions). The outcomes were 3 assessments of 2 efficacy measures: Box and Block Test and Jebsen Taylor Hand Function Test.

Results: A total of 5 children with cerebral palsy (mean 8.4, SD 2.7 years; Gross Motor Function Classification II—2/5, 40%; III—2/5, 40%; and IV—1/5, 20%) and 1 child with obstetric brachial plexus palsy (aged 8 years; Mallet Classification III) received between 8 and 11 sessions of training (10-20 minutes per session), depending on age, motivation, and fatigue. Significant associations were observed between game parameter settings and improvements in motor function, on the one hand, and the type of improvement and disability severity, on the other: with adjusted game parameters goal and time in the range of 70% to 100%, only less affected children improved in the Box and Block Test (+11 blocks vs −1 block), and more affected children improved more in the Jebsen Taylor Hand Function Test (+90 seconds vs +27 seconds).

Conclusions: When defining the difficulty parameters for an exergame, we suggest a classification in levels ranging from very easy to very hard. For practical use, we suggest setting the difficulty for the player to an easy or medium level rather than high-commitment goals, as this leads to a longer playtime with more fun and, therefore, seems to improve the results of the game and, consequently, mobility.

(JMIR Serious Games 2023;11:e36110) doi:10.2196/36110

KEYWORDS
rehabilitation; physical therapy; cerebral palsy; obstetric brachial plexus palsy; serious games; exergames; Kinect
**Introduction**

**Background**

Cerebral palsy (CP) and obstetric brachial plexus palsy (OBPP) are 2 of the pathologies that most frequently cause problems in the upper limbs in children and adolescents. According to the International Classification of Functioning, Disability, and Health (ICF), these problems affect corporal structure and function, limiting, in many cases, activities of daily living [1]. CP is considered a global disorder because of a permanent and nonprogressive lesion in the brain, before its development is complete, producing alterations in tone, posture, and movement, resulting in functional limitations and other associated alterations [2,3]. OBPP is generally caused by an elongation of the brachial plexus during childbirth, although it can also have an intrauterine, genetic, or postural origin. Alterations range from nerve conduction impairment to total rupture of anatomical structures [4,5]. Common problems of both pathologies are muscular atrophy and weakness, imbalance, alteration of the coordination of different muscle groups, secondary joint pathology, and decreased general use of the arms and hands in daily activities [6,7].

Regarding the different conservative treatments for both pathologies, we found that different systematic reviews proposed repetitive, challenging, high-intensity, and motivational interventions [5,6,9,9].

When analyzing the interventions that are considered effective in CP, we found that training with video games or virtual reality (VR) is among the interventions for which there is promising evidence suggesting possible effectiveness [8]. In OBPP, VR associated with other therapeutic strategies is also considered significantly more effective than conventional physical therapy programs [4,10]. Therefore, serious games and VR are considered promising interventions in both conditions.

Systematic reviews that analyzed the effectiveness and clinical utility of VR interventions indicated that motion capture devices have beneficial effects in the rehabilitation of patients with neurological disorders, achieving high levels of compliance, motivation, and commitment that can help improve one or more levels of ICF. The greatest benefits were found in domain 1 of the ICF, being effective in improving upper extremity function, hand coordination, balance function, gait, and postural control. In addition, benefits are associated with aspects such as participation in the community and improvements in psychological and cognitive functions [5,9,11-13].

The benefit of video games is that they offer immediate feedback and rewards that stimulate the brain so that the user wants to play more. Providing challenging and fun environments that keep the player motivated increases the number of repetitions of certain movements compared with conventional therapy. Thus, neuroplastic changes and cortical reorganization, which could produce improvements in performance, are elicited [14,15]. The use of these types of games encourages participants to direct their attention to an external focus, which helps them to learn faster and more accurately than requests from an internal focus [16].

As for motion-controlled video games, most studies use Kinect devices (Microsoft Corp) along with commercial video games, which provide accessibility and portability and are low cost. However, these video games do not address the specific needs of patients who sometimes cannot perform the required movements quickly or efficiently enough to succeed during play. Therefore, different reviews point to the advantages of specifically designed games that provide appropriate challenges and adapt to the player’s speed and quality of movement, thus avoiding boredom and frustration. However, the negative aspect is that commercially available mainstream video games cannot be personalized and that there are still very few offers for therapeutic video games, which are quite expensive [11-13,17-20].

Some commercial solutions we found are the MIRA software platform (MIRA Rehab Limited), [21], VirtualRehab (Evolv) [22], and REHABILITY (Imaginary SL) [23], all of which seem quite complete, but it is not clear from the information provided on the web pages how and to what extent the different exercises involved in the games are configurable for each user.

The common opinion of researchers is that games should complement therapy with both options, accompanied by a professional or used at home [9,12,24]. In any case, it is recommended that computer-based training be followed by the practice of activities in the physical environment to ensure adequate perceptual adjustment and the transfer of the benefits of training to activities of daily living performed in a real environment [24].

**Objectives**

The purpose of this study was to observe how much influence the adjustment of the games’ objectives has on the child’s performance in upper limb movements. We suppose that there is a small range of adjustments that ensures that the challenge for the child is neither too low (leading to boredom) nor too high (leading to frustration). Therefore, we wanted to prove the hypothesis that the optimal settings (of the goal and the time parameters) are those that ensure that the goal can be reached within the given time limit. Furthermore, the aim was to derive some practical adjustment rules from our results that could be useful for our and other similar configurable games.

**Methods**

**Therapeutic Video Game Environment**

The Blexer system has been described in detail in various previous publications [25,26]. It consists of a database and a web platform that integrates different video games and is accessible only to clinicians. On the user side, a PC with a motion capture camera (Kinect X360, Microsoft Corp) is required. The PC hosts one or more games and a middleware [27] that acts as a communicator among the camera, the database, and the games. It provides the games with motion data, downloads the game settings established by the clinician, and uploads the users’ results after each play to the database.

The game used in this study, Phiby’s Adventures (version 1), is a semi-3D third-person video game extensively described in the study by Eckert et al [26]. The main character, Phiby, is a...
small amphibian that must be helped to get out of a valley, which is represented by a modular 2D map with $16 \times 8$ cells. The cells are connected through different obstacles such as rivers, trees, lakes, or trunks that Phiby must pass while finding its way out of the valley. Each obstacle is an independent exercise scene (refer to the description provided in Table 1) that needs arm and trunk movements of different intensities to pass through. The logs obtained from chopping the trunks can be used to build bridges or huts. A hut allows the player to rest and to save the game state for resuming from that point in the next session. With the aim of increasing the player’s interactivity and decision-making, they are given an option to freely choose the path to follow and, therefore, take the obstacles they like. By passing through the cells, the areas of the map are gradually discovered and can be visualized through an arm movement. The 4 exercise scenes have adjustable difficulty parameters (objective and time limit), as explained in Table 1. An initial calibration measures the players’ range of movement for adapting the game to their capacities.

### Table 1. Overview of the 4 exercises inherent in the video game.

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Description</th>
<th>Abilities to train</th>
<th>Pictures</th>
</tr>
</thead>
</table>
| Chop the wood      | Exercise using the affected limb: this exercise provides a specific activity for the use of the affected side in which maximum shoulder flexion is requested based on the initial calibration. The child works on support and shoulder girdle stability to keep the axe elevated until it is loaded as well as on arm speed to lower the axe and split the wood. The amount of wood to split ($n$umber of arm lifts) and the time limit to achieve all can be modified. | - Proximal arm movement (shoulder flexion)  
- Movement speed  
- Initiation of use of assisting hand  
- Reaches  
- Endurance  
- Stabilization of the shoulder and shoulder girdle | ![Image] |
| Dive and eat       | Trunk control exercise: in this exercise, the objective is to catch some of the pieces of plankton floating in the water. The player controls the direction of the fish with their trunk movements. It favors trunk control and can be performed in a sitting or standing position. The number of planktons to catch and the time limit can be modified. | - Trunk control and stability | ![Image] |
| Row the boat       | Symmetrical movement exercise of upper extremities: in this exercise, the symmetrical use of both upper limbs is proposed, and the objective is to row a boat until reaching the opposite bank of the river. The speed of the boat depends on the speed of the user’s movements. The distance to travel ($n$umber of arm movements) and the time limit can be adjusted. | - Proximal movement of the arm  
(shoulder flexion-extension) and elbow  
- Movement speed  
- Initiation of use  
- Hand and arm coordination  
- Execution flow of bimanual activities | ![Image] |
| Climb the tree     | Alternated movement exercise in upper limb abduction: this exercise involves alternating movements of the arms in abduction, and the objective is to reach the top of a tree to see the valley. The parameters to be modified are the height of the tree ($n$umber of arm movements) and the time limit. | - Proximal arm movement (shoulder abduction and flexion)  
- Movement speed  
- Initiation of use  
- Reaches  
- Hand and arm coordination  
- Execution flow of bimanual activities | ![Image] |

### Study Design and Participants

This case study was jointly driven by the Research Center for Software Technologies and Multimedia Systems for Sustainability of the Universidad Politécnica de Madrid (UPM) and the Public School CEIP (Centro de Educación Infantil y Primaria) Pinar de San José in Madrid. UPM developed the Phiby’s Adventures video game and the Blaxter-med web platform used in this study. The game is connected to the platform, which allows for the adjustment of the difficulty of the game parameters and the supervision of the results for each participant.

CEIP Pinar de San José is an integrative primary school (5-12 years) with 28 pupils (out of 950) having motor disabilities (13 CP, 2 OBPP). Informed consent was obtained from the parents and children involved in this study.

Inclusion criteria for the study were the following: being a scholar receiving physiotherapy care at the Pinar de San José school, being aged between 5 and 12 years, having a major functional impairment of an upper limb because of CP or OBPP, and being able to understand and follow the game. Exclusion criterion was poor trunk control that prevents activity from being performed in a freely seated position. The children were classified based on their functional level according to the Gross Motor Function Classification (GMFCS) [28], Manual Ability Classification System (MACS) [28,29], and Mallet Classification [30,31].

### Intervention

Initially, several instructional sessions were conducted for 8 weeks to familiarize the children with the dynamics of the game. Then, the intervention took place for 6 weeks, using the game as a part of the usual 45-minute physiotherapy sessions (2 or 3
times a week), for approximately 10 to 20 minutes, according to age, motivation, and fatigue. In the remaining time, they received the usual physiotherapy treatment. They also continued with activities and therapies outside school.

During the instructional sessions conducted before the intervention, all the children played with the same default difficulty adjustments, which were 60 seconds time limit per exercise with the following goals: chopping 60 trunks, climbing up 100 meters, catching 10 planktons diving, and rowing 100 meters. After these sessions, the objectives of each game were adjusted individually for each child according to our subjective observations during training and with the intention to provide affordable challenges.

During the intervention, physiotherapists constantly supervised the children and verbally motivated them. They were allowed to vary their pose (sitting or standing) depending on their abilities and how they felt that day; however, all participants selected their favorite posture once and then always played in the same way and without any physical help from the therapists. Depending on the results observed, that is, considering whether the objectives were achieved and how quickly or whether the children were bored or stressed, the difficulty of each exercise was readjusted after each session to maintain the challenge. In the case of being tired or unmotivated, the goals were lowered, whereas they were raised when improvement or low challenge was observed. The time limit was kept constant at 60 seconds in all exercises. These decisions were based on subjective observations (of movements and mood) as the objective results (time needed for an exercise and score) were not graphically presented on the web at that moment. They could be revised in form of tables, but this was not useful for making instantaneous objective decisions. All the children were able to complete the activity.

Data Analysis of the Exercise Results and Jebsen Taylor Hand Function Test and Box and Block Test Assessments

The web platform receives the following for each exercise: its difficulty parameters (objective and time limit), the time stamps of entering and leaving, and the score reached. We analyzed the relative scores for each type of exercise, that is, the percentage of movements in relation to the objective established by the therapist, and correlated these data with the outcomes of 3 evaluations: previous evaluation (Vp) 8 weeks before, initial evaluation (Vi) immediately before, and final evaluation (Vf) immediately after the intervention. Between Vp and Vi, conventional therapy was performed, and after Vi, the game was used as a part of the therapy sessions for 6 weeks.

The assessments were made using the Box and Block Test (BBT) [32] (counting the number of blocks put from one part of the box to another; the more blocks, the better the result) and the Jebsen Taylor Hand Function Test (JTHFT) [33] (measuring the time required to perform 7 activities; the quicker, the better the result). Both tests have demonstrated reliability, especially for measuring changes that focus on improving hand dexterity and function in children with CP after intensive intervention. The reliability and responsiveness of these tests are better for short evaluation periods [18,19].

Statistical calculations were performed using SPSS software (version 26; IBM Corp). Descriptive statistics were used to summarize baseline characteristics and feasibility data. We used the Levene test to assess the normality of data distribution. Differences of means were calculated and proved by 2-tailed significance tests, and the differences were considered statistically significant at $P<.05$.

**Ethics Approval**

This study was approved by the Universidad Politécnica de Madrid (UPM) ethics committee and also complied with the Declaration of Helsinki.

**Results**

**Study Design and Participants**

For the study, 54% (15/28) of the children aged 5 to 12 years were screened for eligibility among the ones receiving physical therapy support at the Public School CEIP Pinar de San José in Madrid. Out of these, 47% (7/15) met all inclusion criteria and 7% (1/15) had to be excluded because of poor trunk control and inability to maintain stability during play without external aids; so finally, the analysis group was reduced to 6 children (Table 2). The mean age of the participants was 8.3 (SD 2.4) years; 5 (83%) children with CP were aged 8.4 (SD 2.7) years, and 1 (17%) child with OBPP was aged 8 years. We used the following code to identify the participants: CP/OBPP+age+hand+group.

The participants were divided into 3 groups based on the results of the functional classifications of GMFCS, MACS, and Mallet (Table 2):

1. Group 1: CP (GMFCS ≥III, MACS ≥III; 3/6, 50% participants)
2. Group 2: CP (GMFCS<III, MACS <III; 2/6, 33% participants)
3. Group 3: OBPP (Mallet III; 1/6, 17% participants)

The participants in group 1 had a lower functional level, with an average age of 7 (SD 2.6) years. The participants in groups 2 and 3 presented a similar profile with higher functional levels in both gross and fine motor skills; therefore, these 2 groups were merged into group 2+3 (average age: 9.7, SD 1.5 years). Thus, 2 homogeneous groups of 3 children each were formed, considering similar functionality and age.
Table 2. Details of the participants’ age, sex, diagnosis, affected hand, playing position, motor function ratings, and assigned group.

<table>
<thead>
<tr>
<th>ID</th>
<th>Ageᵃ (years)</th>
<th>Sex</th>
<th>Diagnosis</th>
<th>More affected hand</th>
<th>Play position</th>
<th>Classification</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP5L1</td>
<td>5</td>
<td>Female</td>
<td>CPᵈ</td>
<td>Left</td>
<td>Seated</td>
<td>IV</td>
<td>1</td>
</tr>
<tr>
<td>CP6R1</td>
<td>6</td>
<td>Male</td>
<td>CP</td>
<td>Right</td>
<td>Seated</td>
<td>III</td>
<td>N/A</td>
</tr>
<tr>
<td>CP10R1</td>
<td>10</td>
<td>Male</td>
<td>CP</td>
<td>Right</td>
<td>Standing</td>
<td>III</td>
<td>1</td>
</tr>
<tr>
<td>CP10L2</td>
<td>10</td>
<td>Male</td>
<td>CP</td>
<td>Left</td>
<td>Standing</td>
<td>II</td>
<td>N/A</td>
</tr>
<tr>
<td>CP11R2</td>
<td>11</td>
<td>Female</td>
<td>CP</td>
<td>Right</td>
<td>Standing</td>
<td>II</td>
<td>2</td>
</tr>
<tr>
<td>OBPP8R3</td>
<td>8</td>
<td>Female</td>
<td>OBPP</td>
<td>Right</td>
<td>Standing</td>
<td>N/A</td>
<td>Mallet III</td>
</tr>
</tbody>
</table>

ᵃAge, mean (SD): all=8.3 (2.4) years; CP=8.4 (2.7) years; group 1=7.0 (2.6) years; group 2+3=9.7 (1.5) years.
ᵇGMFCS: Gross Motor Function Classification.
ᶜMACS: Manual Ability Classification System.
ᵈCP: cerebral palsy.
ᵉN/A: not applicable.
ᶠOBPP: obstetric brachial plexus palsy.

Intervention

On average, 9.8 (range 8-11) sessions were conducted per participant, adjusting the duration of each session according to age, motivation, and fatigue. Given the fact that each player needed individual adjustments of the time limit and the objective of an exercise, comparisons could be made only through relative values; therefore, we defined 2 variables that represent the results as percentages of what was achieved in relation to what was configured: \( \text{pct}_{\text{goal}} \) (percentage of goals, equation 1) is the number of movements performed in relation to the objective established by the therapist and \( \text{pct}_{\text{time}} \) (percentage of time, equation 2) is the time needed to finish the exercise in relation to the time limit set.

Table 3 presents the individual results that the participants obtained for each exercise during the intervention sessions. For analyzing how well the target values of the exercises were adjusted in each session according to the child’s abilities, \( \text{pct}_{\text{goal}} \) and \( \text{pct}_{\text{time}} \) were plotted over time. The aim was to avoid excessive exertion and under exertion and to achieve the highest possible level of motivation. Figure 1 shows 4 examples of the performance of different players in different exercises, comparing the time and score achieved during the intervention.

Participant CP5L1 achieved a 100% score on the rowing exercise (Figure 1, top left) in less than half the time, and our observation was that the child quickly became bored. The plot on the right side of the top row in Figure 1 shows the opposite situation for the same child during the dive exercise: 100% of the time limit was almost always used but to no avail, and we also observed difficulties and frustration. However, in this case, it was not possible to set an easier target because it was already set at the minimum value. The third case (CP6R1 in the climbing exercise) shows a good balance between the score obtained and time taken, with both percentages being similar and mostly close to 100%. The child also played with fun and seemed challenged. The fourth case, CP10L2 in the chopping exercise (Figure 1, below right), shows a readjustment situation: the first 16 exercises were completed with a high goal and a short time limit so that the goal was almost never reached in the given time. We observed stress and reduced the goal (keeping the time limit the same), resulting in a setting that allowed reaching the goal 100% and in a short time.

The graphical representation of all individual results from Table 3 is shown in Table 4 and in the graph in Figure 2. It shows a group of settings in the top right where both the score and the time spent are high (between 70% and 100%). Very few cases had settings that resulted in very low scores or high scores obtained in a short amount of time (blue and yellow outliers). Accordingly, and also considering our observations of the children during play, we classified the setting ranges from very easy to very hard. To do this, we divided each axis into thirds.
Table 3. Individual playing results, relative to the adjustments.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Chop</th>
<th>Row</th>
<th>Dive</th>
<th>Climb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$pct_{goal}$ (%)</td>
<td>$pct_{time}$ (%)</td>
<td>$pct_{goal}$ (%)</td>
<td>$pct_{time}$ (%)</td>
</tr>
<tr>
<td><strong>Group 1, value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP5L1</td>
<td>62</td>
<td>94</td>
<td>100</td>
<td>36</td>
</tr>
<tr>
<td>CP6R1</td>
<td>74</td>
<td>94</td>
<td>92</td>
<td>79</td>
</tr>
<tr>
<td>CP10R1</td>
<td>75</td>
<td>97</td>
<td>82</td>
<td>73</td>
</tr>
<tr>
<td>Group 1, mean (SD)</td>
<td>70 (6)</td>
<td>95 (1)</td>
<td>91 (7)</td>
<td>63 (19)</td>
</tr>
<tr>
<td><strong>Group 2+3, value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP10L2</td>
<td>98</td>
<td>52</td>
<td>95</td>
<td>75</td>
</tr>
<tr>
<td>CP11R2</td>
<td>93</td>
<td>59</td>
<td>92</td>
<td>67</td>
</tr>
<tr>
<td>OBPP8R3</td>
<td>89</td>
<td>85</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>Group 2+3, mean (SD)</td>
<td>93 (4)</td>
<td>66 (14)</td>
<td>88 (8)</td>
<td>72 (4)</td>
</tr>
<tr>
<td>All participants, mean (SD)</td>
<td>82 (13)</td>
<td>80 (18)</td>
<td>90 (8)</td>
<td>67 (14)</td>
</tr>
</tbody>
</table>

a $pct_{goal}$: percentage of points obtained with respect to those established.
b $pct_{time}$: required time percentage of set time limit.

Figure 1. Four examples of different player experiences. Top left: easy play (100% goal achieved in a very short time). Top right: difficult play (100% time consumed but nearly no hits). Bottom left: challenged play (high percentage of goals, high time consumption). Bottom right: initially difficult play, finally easy play. The following code was used to identify the participants: CP/OBPP+age+hand+group.

Table 4. Relation between playing time, score, and adjusted game difficulty level.

<table>
<thead>
<tr>
<th>$pct_{time}$ (%)</th>
<th>$pct_{goal}$ (%)</th>
<th>Game difficulty level</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤30%</td>
<td>&gt;$pct_{time}$</td>
<td>Very easy</td>
</tr>
<tr>
<td>&gt;30% and ≤70%</td>
<td>&gt;$pct_{time}$</td>
<td>Easy</td>
</tr>
<tr>
<td>&gt;70% and ≤100%</td>
<td>&gt;70% and ≤100%</td>
<td>Medium</td>
</tr>
<tr>
<td>&gt;$pct_{goal}$</td>
<td>&gt;30% and ≤70%</td>
<td>Hard</td>
</tr>
<tr>
<td>&gt;$pct_{goal}$</td>
<td>≤30%</td>
<td>Very hard</td>
</tr>
</tbody>
</table>

a $pct_{goal}$: percentage of points obtained with respect to those established.
b $pct_{time}$: required time percentage of set time limit.
**Data Analysis of the Exercise Results and JTHFT and BBT Assessments**

Table 5 presents the data of the individual results of the 3 evaluations, and Figure 3 is a visualization of these values showing the improvement in both tests.

In addition, the graph in Figure 4 has been created to perform a visual analysis of the relation of the JTHFT ratings with the BBT scores for each participant. We added a hull around the 3 evaluations to show at first glance how much a child improved between assessments: the larger the shape, the more the results differed between the JTHFT and BBT. A very horizontal slant indicates better performance in the BBT, and a vertical slant indicates more improvement in the JTHFT. Round shapes indicate few differences between both assessments, and a slant toward 45° indicates that the improvements were similar.

<table>
<thead>
<tr>
<th>Participants</th>
<th>JTHFT (seconds)</th>
<th>BBT (number of blocks)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V_p^a$</td>
<td>$V_i^b$</td>
</tr>
<tr>
<td><strong>Group 1, value</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP5L1</td>
<td>473.1</td>
<td>526.6</td>
</tr>
<tr>
<td>CP6R1</td>
<td>719.5</td>
<td>627.5</td>
</tr>
<tr>
<td>CP10R1</td>
<td>949.4</td>
<td>732.7</td>
</tr>
<tr>
<td>Group 1, mean (SD)</td>
<td>714.0</td>
<td>(194.49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group 2+3, value</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP10L2</td>
<td>157.4</td>
<td>162.1</td>
</tr>
<tr>
<td>CP11R2</td>
<td>206.3</td>
<td>164.4</td>
</tr>
<tr>
<td>OBPP8R3</td>
<td>65.9</td>
<td>58.4</td>
</tr>
<tr>
<td>Group 2+3, mean (SD)</td>
<td>143.2</td>
<td>(58.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants, mean (SD)</td>
<td>428.6</td>
<td>(349.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between groups</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference of means</td>
<td>−570.83</td>
<td>−500.6</td>
</tr>
<tr>
<td>P value</td>
<td>.02</td>
<td>.002</td>
</tr>
</tbody>
</table>

$^a$ $V_p$: previous evaluation.

$^b$ $V_i$: initial evaluation.

$^c$ $V_f$: final evaluation.

In addition, we assembled the evaluation data with exponential trend lines to demonstrate the correlation between them. The resulting curves from the 2 initial assessments did not differ, indicating that the children’s mean scores were very similar.
However, the curve of the third assessment stood out significantly, showing a more vertical slope, indicating that there were greater differences between JTHFT and BBT in this final assessment, pointing to an improvement in skills owing to the games.

Figure 3. Visual comparison between the 3 evaluations with the Jebsen Taylor Hand Function Test (JTHFT) and Box and Block Test (BBT). Vf: final evaluation; Vi: initial evaluation; Vp: previous evaluation.

Figure 4. Visualization of the dependencies between evaluations, test types, and groups. Within parentheses, the percentage of time played with easy and medium adjustment of the game parameters is given. The following code was used to identify the participants: CP/OBPP+age+hand+group.

Discussion

The main finding of this study is the categorization of the parameter adjustments into 5 levels of difficulty, which may be of high interest to a therapist who has to decide on the configuration of different exercises. Here, the easy and medium levels seem to be the most effective, as the final assessment tests show greater improvements for those children who played longer in these adjustment ranges.

Participants

The number of children in this study seems small at first glance, but given our conditions (being a common primary school that integrates 28 children with very different types of physical disabilities), we were fortunate to find 6 children of similar conditions and age and that they were allowed to participate in this study. However, their physical abilities were different because of different degrees of affection, so we decided to analyze them in 2 homogeneous subgroups of the same size, which turned out to be a positive decision, as can be observed from the results.

Intervention

Table 3 presents the results of play for each child and each type of exercise as relative values between the scores and the adjustments. In this way, it was possible to quickly observe in which cases the parameter settings led to better results (the higher the percentage, the more appropriate the parameter adjustment was to the child’s capacity). The graphical representation of these values (refer to the graph in Table 4) led to a distinction of 5 levels of difficulty ranging from very easy to very hard and reveals that most exercises were played with settings that led to achievements of 70% to 100% for both values (upper right part of the graph in Table 4). This means that the objectives were achieved in most cases and that the time given was sufficient, neither too short nor too long. We also observed that in those cases, the children played with more fun and enthusiasm. In the other ranges, they performed worse and felt frustrated or bored. When the time limit was not considered a challenge, as the highest score could be achieved without much effort, the setup seemed too easy. When the goal of an exercise could not be achieved within the given time limit, the participants became stressed and frustrated, so adjustments were probably too difficult. These observations led us to consider the
range between 70% and 100% as a medium difficulty level, which could be the optimal range of adjustment.

Figure 1 visualizes 4 typical cases showing cases of play that were too easy, too hard, and just right. Participant CP5L1 achieved a 100% score in less than half of the given time limit, so this could be considered a very easy adjustment. In the diving exercise, the same child needed the full amount of time given but was unsuccessful, so the configuration appeared to be very hard. CP6R1 achieved the score within the given time, so the configuration of this exercise was considered well suited for this child; we would call it an optimal setup. CP10L2, in the beginning, required the entire time given and achieved very low scores. After readjustment, the child never failed and quickly achieved the goal. Both situations were not ideal, too difficult at first and too easy afterward.

During the intervention, adjustments were made to the parameters of the game based on subjective observations because at that time, the therapist did not have a graphic display of the results available on the web. The adjustments made were expected to be optimal; however, subsequent analysis showed that not all children received an adequate adjustment to their abilities. For example, we noticed an incorrect setting for the participant CP5L1, both in the training sessions and in the intervention sessions, which could possibly have been avoided. Therefore, we consider it crucial that a therapist has an objective measure to verify whether a game setting is well adjusted for each participant. We are convinced that if we had had a tool that showed the range in which a child was playing, our adjustments would have been made more quickly and accurately.

In addition to the general results, it is worth mentioning some differences among the 4 types of exercises (or minigames) inherent in the game. Of them, 3 (75%) are aimed at arm movements (rowing, climbing, and chopping), and the child can fully concentrate on them because the avatar does not change its position when the child's position changes. Instead, the dive exercise is fully controlled in 3D; that is, each corporal movement translates into a change in the position of the avatar in the virtual space. This exercise was created to train trunk stability; however, it turned out to be the most difficult to control, especially for the youngest and most affected children. Therefore, this exercise produced frustration in some cases, especially for CP5L1. Here, we set the goals to low values, because we noticed that children avoided this exercise. They liked the chopping exercise best, which has a great animation with a sparking axe and logs falling to the sides. Climbing and rowing were the 2 easiest exercises, and the children got much faster after learning how to do them.

Data Analysis of the Exercise Results and JTHFT and BBT Assessments

The results obtained in the 3 evaluations of the JTHFT and BBT showed clear improvements for almost all children, especially between $V_i$ and $V_j$ (before and after the intervention with the video game): a reduction in time for the JTHFT and an increase in the number of blocks for the BBT (refer to the column $V_j - V_i$ in Table 5).

The results of group 1, which includes children with lower capacities, clearly differ from those of group 2+3, which can be observed in the difference in mean values for both groups. Group 1 had longer times for the Jebsen tests and moved fewer blocks in the BBT (3 to 5 times fewer blocks). The difference in means can be considered significant despite the small sample size; after checking for normality through Levene test, all $P$ values were <.05 for the 3 assessments (JTHFT: $V_p$, $P = .02$; $V_i$, $P = .002$; and $V_t$, $P = .002$; BBT: $V_p$, $P = .03$; $V_i$, $P = .04$; and $V_t$, $P = .01$).

However, the improvements in both tests achieved through the intervention were different: group 1 improved their results in the JTHFT more than group 2+3. In $V_p$, the average times achieved were about 5 times longer, but the improvement compared with $V_i$ was 89.1 seconds, which was 3 times better than the improvement of group 2+3 (26.9 s). For the BBT, this was the other way around: after the intervention, group 2+3 achieved moving 11 blocks more than those moved before the intervention (+13%), whereas group 1 did not show any improvement caused by the game or related to the game.

An important correlation can be observed when contrasting the data obtained from the BBT and JTHFT evaluations with the time they played in the easy or medium difficulty range. As can be observed in Figure 3, the children who played most of the time with this configuration (refer to the percentages within parentheses) were the ones who achieved the best evaluations in either the BBT or JTHFT.

Looking at the individual cases, it can be observed that the improvement between the initial and final BBT ratings is best for CP11R2 (who played 69% of the time with the easy or medium settings) and very high for OBPP8R3 (who played 51% of the time on easy or medium). CP10L2 played 61% of the time at those ranges and improved similarly in the BBT and JTHFT. CP10R1 (42%) and CP6R1 (48%) improved more in the JTHFT. Finally, CP5L1 unfortunately had very hard settings for her abilities, and as a result, she was the only child who did not improve, neither in the BBT nor in the JTHFT. As explained before, the reason was probably the young age and the general difficulty of the diving exercise, which was already set to the minimum possible goal for this child.

From these results, we understand that a correct adjustment range generates a certain challenge that leads to a longer playing time and highly motivates the player to reach the goal (a fact frequently mentioned in the literature and also called “flow” [34,35]). This is optimal from the therapeutic point of view, as overexertion and underload are avoided while achieving the highest possible level of motivation.

Comparing our proposed level adjustment in Table 4 with the generally published guidelines for game level adjustment, we observe a great coincidence. For example, Sepulveda et al [34] presented the ratio between player performance ($C_p$) and expected performance ($E_p$), as a method to measure the general performance or quality of adjustment:

$$\text{Performance} = \frac{C_p}{E_p} \ (3)$$
A value next to 1 means that the challenge is appropriate for the player, and with an error range set to 0.2, a value <0.8 indicates the game is too difficult and a value >1.2 means the game is too easy. If we apply this formula to our data, the relation \( \frac{\text{pct}_\text{goal}}{\text{pct}_\text{range}} \) would be equivalent to \( Cp \), with \( Ep=1 \), the ratio expected and desired by the therapist. With this, our very easy level can reach up to 3.3, easy up to 2.3, medium between 1.4 and 0.7, hard down to 0.4, and very hard down to 0.3. Therefore, compared with Sepulveda et al [34], we apply a larger and asymmetrical error margin, which is 0.3 for the harder adjustments and 0.4 for the easier adjustments. This is also graphically confirmed in the figure in Table 4, where most settings fall on the harder side. It has to be mentioned here that this classification should be confirmed or refined in more extensive follow-up studies.

Many publications can be found in the area of automatic and dynamic difficulty adjustment of game levels, as reviewed by Zohaib and Hideyuki [35]. It would be interesting to apply our ranges to one of the proposed algorithms and test whether they could be useful for an automatic adjustment of exercise games for people with disabilities. Furthermore, to the best of our knowledge, there are no publications on the adjustment of the level of play for people with motor disabilities; most approaches consider “accessibility,” for example, the studies by Bierre et al [36], Cairns et al [37], and Carr et al [38], but this is not the same. Accessibility deals with different disabilities and adapts interfaces. Making a game “easy” does not necessarily guarantee that people with disabilities can play it. Game level adjustment goes beyond accessibility, as an accessible game still needs to adapt to the players’ abilities like any other game to challenge the player; only the rules should not be the same as those for abled people.

Limitations
As stated before, on some occasions, the Kinect has difficulty capturing the smallest students, and specific problems were detected in handling the main character of the game, which led to some frustration. We are currently developing a new version of the game based on Kinect One, which has much better detection possibilities.

Furthermore, it is necessary to perform tests with larger population groups and analyze the changes that would appear at a longer intervention time. For future studies, it is important to assess whether benefits are maintained in the long term and analyze motivation. Currently, we are conducting a long-term study on the same game but using an improved web interface that visualizes the range in which the users are playing. In this way, readjustments can be made according to an objective measure and, if necessary, after each session.

Conclusions
The use of video games is considered a viable intervention to carry out in physical therapy sessions in the school environment. We observed that when the parameters were not well adjusted, the play time was short, which led to less changes in motor function (proved with the BBT and JTHFT evaluations).

The adjustment of the game seems to be related to the motor benefits found, so it would be appropriate to have access to the patients’ game performance data easily and continuously to be able to modify the parameters in time and to work within the adjusted range. Therefore, we consider it a good option to use the criteria established in this study to achieve an adequate adjustment, seeking that the child performs most of the game in the range between 70% and 100% of the goal and time. However, exhaustive tests must be performed now to prove these values.

Acknowledgments
This work was supported in part by the Spanish and European Social Funding Committee provided through the Europe 2020 program for the employment of young people (Reference Number PEJD-2017-PRE/TIC-4825). The funder did not play a role in the design, conduct, or reporting of this study. The authors thank all children for their participation and their parents for their consent.

Conflicts of Interest
None declared.

References


Abbreviations

BBT: Box and Block Test
CEIP: Centro de Educación Infantil y Primaria
CP: cerebral palsy
GMFCS: Gross Motor Function Classification
ICF: International Classification of Functioning, Disability, and Health
JTHFT: Jebsen Taylor Hand Function Test
MACS: Manual Ability Classification System
OBPP: obstetric brachial plexus palsy
UPM: Universidad Politécnica de Madrid
VI: final evaluation
Vi: initial evaluation
Vp: previous evaluation
VR: virtual reality
Negami: An Augmented Reality App for the Treatment of Spatial Neglect After Stroke

Britta Stammler¹; Kathrin Flammer²; Thomas Schuster³; Marian Lambert³; Hans-Otto Karnath¹

¹Center of Neurology, Division of Neuropsychology, University of Tübingen, Tübingen, Germany
²Flammer & Gläser UXplain GbR, Karlsruhe, Germany
³XPACE GmbH, Pforzheim, Germany

Corresponding Author:
Hans-Otto Karnath
Center of Neurology
Division of Neuropsychology
University of Tübingen
Hoppe-Seyler-Strasse 3
Tübingen, 72076
Germany
Phone: 49 7071 29 80467
Email: Karnath@uni-tuebingen.de

Abstract

Background: A widely applied and effective rehabilitation method for patients experiencing spatial neglect after a stroke is “visual exploration training.” Patients improve their ipsilesional bias of attention and orientation by training exploration movements and search strategies toward the contralesional side of space. In this context, gamification can have a positive influence on motivation for treatment and thus on the success of treatment. In contrast to virtual reality applications, treatment enhancements through augmented reality (AR) have not yet been investigated, although they offer some advantages over virtual reality.

Objective: This study aimed to develop an AR-based app (Negami) for the treatment of spatial neglect that combines visual exploration training with active, contralesionally oriented rotation of the eyes, head, and trunk.

Methods: The app inserts a virtual element (origami bird) into the real space surrounding the patient, which the patient explores with the camera of a tablet. Subjective reports from healthy elderly participants (n=10) and patients with spatial neglect after stroke (n=10) who trained with the new Negami app were analyzed. Usability, side effects, and game experience were assessed by various questionnaires.

Results: Training at the highest defined difficulty level was perceived as differently challenging but not as frustrating by the group of healthy elderly participants. The app was rated with high usability, hardly any side effects, high motivation, and entertainment. The group of patients with spatial neglect after stroke consistently evaluated the app positively on the dimensions of motivation, satisfaction, and fun.

Conclusions: The Negami app represents a promising extension by adding AR to traditional exploration training for spatial neglect. Through participants’ natural interaction with the physical surrounding environment during playful tasks, side effects as symptoms of cybersickness are minimized and patients’ motivation appeared to markedly increase. The use of AR in cognitive rehabilitation programs and the treatment of spatial neglect seems promising and should receive further investigation.

(JMIR Serious Games 2023;11:e40651) doi:10.2196/40651

KEYWORDS
spatial neglect; gamification; augmented reality; visual exploration training; stroke rehabilitation; serious games; rehabilitation; stroke

Introduction

Spatial neglect is the dominant disorder after a brain lesion of the human right hemisphere [1,2]. The most common cause of spatial neglect is a large stroke in the right middle cerebral artery area that affects a perisylvian network consisting of the superior/middle temporal cortex and the parietal and ventrolateral frontal cortices [3]. Affected patients behave as
though the left side of their surrounding space no longer exists. The patient’s eyes and head are sustainedly oriented to the side of the brain lesion (ie, mostly to the right side) [4,5]. When searching for objects, the patient’s visual and tactile exploration is shifted to the right side. Accordingly, objects or people located on the left side are ignored and not found. Many therapeutic approaches to spatial neglect thus entail exercises (eg, search training on large projection screens) and activities (eg, reading and copying tasks, picture descriptions) that require increased and active turning of the patients toward the neglected contralesional side [6-9]. In this training, termed “visual exploration training” or “visual scanning training,” exploration movements are improved, and compensatory search strategies are practiced, leading to improvements in neglect behavior in important everyday situations.

A critical component of neurorehabilitation and its success is related to patients’ motivation to train [10]. Improving the latter is one of the main aspects of recent rehabilitation efforts known as “gamification” [11,12]. Gamification describes the application of game design elements in a nongame context to motivate desired behavior. Treatment motivation in stroke rehabilitation can additionally be increased by gamified feedback (ie, giving feedback about the performance in the corresponding task by awarding points) [13].

The principle of gamification is further emphasized when combined with virtual reality (VR) methods. The principle of active exploration training in spatial neglect has already been extended by several VR approaches. Nonimmersive VR procedures, in which a virtual environment is displayed on a monitor with which the patient interacts without being immersed in it, have been successfully used in group studies for therapeutic purposes [14-16]. On the other hand, immersive VR procedures, in which the patient becomes part of a virtual 3D environment surrounding them by means of VR goggles, have also been developed, gearing toward active exploration scenarios. Recent proof-of-concept studies have shown that healthy participants and patients who have had a stroke perceive this technique as highly motivating and enjoyable [17-19]. Moreover, immersive VR applications have already been tested with individual patients with spatial neglect for treatment purposes [20,21].

Another promising way to increase training motivation in poststroke rehabilitation is through augmented reality (AR). AR describes the computer-generated addition of virtual elements to reality. The visual, real world is augmented by virtual images and figures via a video camera of an electronic device such as a tablet. The virtual addition takes place in real time in the sense of a live transmission and thus combines the real and the virtual [22]. Bakker and colleagues [23] developed the first game combining AR elements with visual scanning training in the context of neglect rehabilitation. It consisted of participants searching for virtual images projected onto a wall in the environment of a museum. Design decisions were made based on expert opinion as well as the opinion of patients with spatial neglect. Features to increase extrinsic motivation were rated as the most important design decision.

In this development and feasibility study, the application of an AR technique for the treatment of spatial neglect is further developed. It is based on the observation that visual exploration training seems to be particularly effective when combined with an active rotation of the trunk contralesionally [24]. An attractive extension of this treatment principle is the innovative app named Negami. This app is based on the principle that patients are playfully motivated to orient themselves to their neglected side of the real room by (1) following and (2) searching for a virtual element (in this case an origami bird), actively exploring space by turning their gaze, head, and trunk. In developing the game, special consideration was given to the age group to ensure high usability. Our goal was to develop a task that would motivate the patients and that they would enjoy. By eliciting the subjective views of both healthy elderly participants (aged 57 to 89 years) and patients with spatial neglect after stroke regarding our AR-assisted exploratory therapy app, we also aimed to uncover barriers to its use and make recommendations to improve the future development of therapies with AR.

Methods

Participants

Ten elderly neurologically healthy individuals with a mean age of 66 (SD 11.9) years, including 5 (50%) females, participated in this study. In addition, we included 10 patients, including 2 (20%) females, who were experiencing spatial neglect and left-sided hemiparesis after a stroke in the neurological right hemisphere —5 (50%) with infarction and 5 (50%) with hemorrhages. They had a mean age of 61.3 (SD 15.2) years and a poststroke interval of 138.4 (SD 192.1) days when they were tested for spatial neglect. More details of the patients are provided in Multimedia Appendix 1.

The inclusion criterion for the patients was the presence of spatial neglect after a stroke. In addition to clinical behavioral observations, diagnostic criteria had to be met in at least 2 of the following 4 neglect tests: the Letter Cancellation Test [25], the Bells Test [26], a Copying Task [27], and a Line Bisection Task [28]. All 4 neglect tests were performed on a Samsung S7+ tablet with screen dimensions of 285 × 185 mm. The severity of spatial neglect in the cancellation tasks was determined by calculating the center of gravity of the target stimuli marked in the search fields, known as the Center of Cancellation (CoC). A CoC value ≥0.08 indicated left-sided spatial neglect [3,29]. The Copying Task consisted of a complex scene consisting of 4 objects (fence, car, house, and tree), wherein points were assigned based on missing details or whole objects. One point was given for a missing detail and 2 for a whole object. The maximum number of points was therefore 8. A score higher than 1 (ie, >12.5% omissions) indicated neglect [27]. In the Line Bisection Task [28], patients were presented with 4 different line lengths eight times each (ie, 32 lines in total). The cut-off value for spatial neglect was an end point weightings bias (EWB) value ≥0.07 [30].

In the group of patients, the mean CoC on the Letter Cancellation Test was 0.42 (SD 0.28), the mean CoC on the Bells Cancellation Task was 0.38 (SD 0.29), the score on the copying test was 4.4 (SD 1.51), and the EWB score on the Line Bisection Task was 0.33 (SD 2.5).
Ethics Approval
All participants gave their informed consent to participate in the study, which was conducted following the ethical standards of the 1964 Declaration of Helsinki. This study was also approved by the ethics committee of the University Clinic of Tübingen (373/2021B02).

The Negami App

Overview
Negami [31] is a mobile app that enables the treatment of spatial neglect using gamified AR. It is optimized for use on tablets but also works on smartphones. In this study, a 3rd generation Apple iPad Pro 12.9” was used to run the app and conduct the therapy sessions. Negami provides 2 modes (Task A and Task B), which are detailed in the subsequent section. In both tasks, the live video stream from the devices’ back cameras is displayed on the screen. Additionally, a virtual element (origami bird) is added (augmented) as a 3D object that must be followed (Task A) or searched for (Task B) by moving the tablet in 3D space. An overlaying crosshair user interface (UI) element provides visual feedback and guidance for the patient.

Technical Background
The Negami app can be achieved via the Negami website. The app was implemented in C# programming language using the open-source cross-platform framework Xamarin (Microsoft Corp) [32]. It allows the creation of native Android, iOS, and Windows apps by sharing UI and code across platforms. The Xamarin platform consists of Xamarin.iOS and Xamarin.Android projects, which are implementations of the Mono framework for iOS and Android, respectively. Mono is a cross-platform open-source .NET compatible software framework. Additionally, Xamarin includes the Xamarin.Forms library for the creation of cross-platform user interfaces using XAML (Extensible Application Markup Language). A UI created in Xamarin.Forms is automatically mapped to native controls on each platform during the build process. Xamarin was chosen for the development of the Negami app because it provides native runtime performance, which is important for game-based apps, as well as an abstraction layer over many platform-specific application programming interfaces (APIs), thus increasing development speed. Additionally, Xamarin offers a set of bindings for full access to all device-specific APIs (Figure 1). Apps built with Xamarin are compiled into native packages for each platform. For the Android platform, the written code is compiled down to Intermediate Language (IL), which is then just-in-time compiled to native assembly on app launch. On iOS, the app is fully ahead-of-time compiled into native ARM (Advanced Reduced Instruction Set Computer Machine) assembly code (Figure 1 [33]).

Figure 1. Architecture of an application developed using Xamarin.iOS (left) and Xamarin.Android (right) [33]. AOT: ahead of time; API: application programming interface.

Utilizing Xamarin, the implementation of the Negami app’s entire business logic and user interface could be shared across platforms. The development followed the Model-View-View-Model design pattern and applied other best practices such as the use of an inversion of control container for dependency injection, commands, and event aggregators. Additionally, the app’s functionality was verified using both unit tests and manual testing. To facilitate the data collection process, a data synchronization feature was implemented that regularly syncs the local SQLite database (which enables the app’s offline usage) containing metadata about each patient and app settings with a remote Microsoft SQL (Structured Query Language) database utilizing the Dotmim.Sync library. As the actual data for each completed exercise session were too large to be stored efficiently in a database, a file-based synchronization was implemented using the Azure Blob Storage (Microsoft Corp) service.

Unlike the business logic and UI elements, the AR aspects of the app had to be implemented in a platform-specific way because Xamarin provides no abstraction layer for these APIs. The implementation was only completed for the iOS platform. For this, the Apple ARKit platform [34] was used (through...
native bindings provided by Xamarin.iOS) to implement all AR-related logic, such as world tracking, placing and rendering of the objects in 3D space, and retrieving the current screen position of the bird. For Android, the Google ARCore library was utilized in the future to implement the same logic.

**Design**

Since the target group consists mainly of elderly people, the design of the app was tailored to their needs to facilitate the most positive user experience. The app was intended to reduce the distance and strangeness that this user group often feels from digital devices. We did not want it to feel like a monotonous, mandatory clinical treatment. Instead, we wanted patients to playfully find the motivation to work on their recovery.

A further aim was to ensure that patients—after discharge from inpatient treatment—can safely use the app to practice on their own or with family members to make further progress. This is achieved through short distances in operation and simple instructions and assistance. The same applies to large control buttons that are easy to reach even with shaky or stiff fingers. We also took into account that patients with spatial neglect ignore information on the contralesional side. For example, this was considered when placing and labeling the exercise buttons. Simple letters like “A” and “B” presented on the right tablet side were used instead of complex icons or text. Moreover, we took into consideration that patients with spatial neglect are often hemiparetic. Instead of holding the tablet from the right and left with both hands, hemiparetic patients can hold it with the unaffected hand from the bottom/center. However, if the tablet is too heavy for the patient to hold in this single-handed position, a loop with Velcro can be attached to the patient’s wrist, which, in turn, is attached to the middle of the back of the tablet to help the patient carry the weight.

In addition to the patients, there is another user group: the therapists. This group of users has different requirements for the app than the patients. Since therapists are often responsible for several patients concurrently, they need to treat patients as efficiently as possible. For this purpose, the app offers the therapist the possibility to easily adjust the corresponding performance level of each patient to their current treatment status. The development of each patient’s skills is easily visible to the therapist. Further, the app offers the possibility to create a plan for independent therapy at home. The patient’s progress thus is recorded and can be viewed by the therapist at any time, allowing them to make the exercises more difficult or easier depending on the patient’s level of ability.

**Tasks Provided by the Negami App**

During the typical use of Negami, the participant sits in a chair or a wheelchair. If possible, the tablet is held with both the left and right hands. If the participant is hemiparetic, the tablet can be held from below with one hand positioned in the middle. Using arm movements in combination with trunk rotations, the tablet can then be moved to the left or right through real space (Figure 2). The participants perform 2 different tasks in succession.

*Figure 2.* Patient performing tasks of the Negami app. A virtual element (origami bird) was added (augmented) to the video stream of the surroundings produced by the camera of a tablet.
The patient’s first task (Task A: “follow the bird”) is to follow the virtual origami bird through real space. The patient’s straight-ahead eye/head/body orientation is set as 0. Beginning at set point 0, the bird flies with random sinusoidal movements toward one side of space (in neglect patients toward the contralesional, neglected side). While doing so, the patient sees an orange circle in the center of the screen (Figure 3). The bird should be held in this circle while performing the task. As soon as the patient fails to successfully follow the bird and keep it in the circle during its flight, they are provided with an additional orientation aid. A blue compass needle appears showing the patient in which direction the bird is located (see Figure 3). The task is successfully completed as soon as the bird has finished its trajectory and the bird has been positioned centrally in the orange circle. During the performance, the patient receives auditory feedback. Every time the patient manages to get the bird into the orange circle, a short, bright tone is presented. If the patient manages to keep the bird continuously in the orange circle, the tone is presented every 2 seconds. When the task has been successfully completed, the patient receives feedback auditorily again through a different tone and visually through the change of the color of the circle to green.

At the end of a successful trial, the options “change angle” and “lead to point 0” are provided on the screen. If the patient/therapist wants to repeat the task, they can select “lead to point 0,” and if the difficulty of the task must be varied, the patient/therapist can select “change angle.” The patient/therapist has 3 different difficulty levels to choose from: easy, medium, or difficult (Table 1). The difficulty levels are saved as templates.

Figure 3. Task A: “follow the bird.” The patient has the task to follow the flying origami bird and to keep the bird within the orange/blue circle (A). If the task is successfully solved the circle turns green (B).

Table 1. Difficulty levels provided with Task A (“follow the bird”).

<table>
<thead>
<tr>
<th>Level</th>
<th>Maximum distance of the bird’s path</th>
<th>Speed of bird movement</th>
<th>Amplitude of bird movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>up to ±35°</td>
<td>2.3°/sec</td>
<td>6.8°</td>
</tr>
<tr>
<td>Medium</td>
<td>up to ±55°</td>
<td>4.1°/sec</td>
<td>8.0°</td>
</tr>
<tr>
<td>Difficult</td>
<td>up to ±85°</td>
<td>8.1°/sec</td>
<td>11.3°</td>
</tr>
</tbody>
</table>

aAngular degrees are participants’ space coordinates, starting from their straight-ahead eye/head/body orientation, which is set as 0.
bsec: second.

In the second task of the Negami app (Task B: “find the bird”), the therapist hides the virtual origami bird in the room surrounding the patient. For this purpose, the patient’s straight-ahead eye/head/body orientation again is set as 0. Starting from this location, the therapist hides the bird somewhere on the left/right side of the patient without the patient seeing it. The best way to do this is for the therapist to step behind the patient and hold the tablet above the patient’s head in such a way that they cannot see where exactly the tablet is pointing. The area within the bird that can be hidden by the therapist is predefined by the app depending on the chosen difficulty level of the task (Table 2). Once the therapist has positioned the bird in space by clicking on the center of the screen, they return the tablet to the patient’s hand. The patient is then instructed to find the bird, and once found, to position it centrally into the orange circle. No time limit is given. Should the patient show difficulties in finding the bird, it is possible to provide the patient with orientation assistance by turning on the blue compass needle (Figure 4). The task is successfully solved when the patient finds the bird and places it in the center of the circle. As with Task A, the circle then turns green (Figure 3), and the patient hears a bright tone signaling the successfully solved task. After successful completion of the task, options “change angle” and “lead to point 0” are provided (see Figure 3) so that the task can be repeated or changed in difficulty.
Table 2. Difficulty levels provided with Task B (“find the bird”).

<table>
<thead>
<tr>
<th>Level</th>
<th>Extension of the area (along the horizontal dimension of the surroundings) within which the bird can be hidden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>0° to −40° (or 0° to +40°)</td>
</tr>
<tr>
<td>Medium</td>
<td>−40° to −75° (or +40° to +75°)</td>
</tr>
<tr>
<td>Difficult</td>
<td>−75° to −90° (or +75° to +90°)</td>
</tr>
</tbody>
</table>

Angular degrees are the participants’ space coordinates, starting from their straight-ahead eye/head/body orientation, which is set as 0. Negative values indicate positions to the participant’s left, while positive values indicate positions to the participant’s right.

Procedure
All neurological patients received training with Negami in their room in the ward of the respective rehabilitation facility. The healthy participants also received the training indoors for comparability. The healthy participants had 1 training session with Negami that lasted 20 to 25 minutes. They completed Task A in the first half and Task B in the second half, each at the highest difficulty level. If patients were not able to use both hands to hold the tablet due to hemiparesis, they were instructed to hold the tablet with the unimpaired hand only by grasping the tablet in the middle from the top or bottom. A video example of a patient working with Negami is shown in Multimedia Appendix 2. To assess feasibility, usability, and game experience, the System Usability Scale (SUS) [35], Simulator Sickness Questionnaire (SSQ) [36], and Perception of Game Training Questionnaire (PGTQ) [37] were administered to the participants directly after the training session.

To assess usability, the participants answered the 10 questions of the SUS through a 5-point Likert scale ranging from 1 (fully disagree) to 5 (fully agree). The mean score was calculated for each participant across all questions. To capture the presence of side effects of simulation technologies, the SSQ was applied. For this purpose, the presence of a series of symptoms grouped under “cybersickness” was tested by a 4-point Likert scale ranging from 0 to 3 (not at all, slightly, moderately, and strongly). The following 15 symptoms were recorded in total: general discomfort, fatigue, headache, eyestrain, difficulty focusing, increased salivation, sweating, nausea, difficulty concentrating, fullness of the head, blurred vision, dizziness, vertigo, stomach awareness, and burping. Again, the mean score was calculated across all items for each patient. Finally, the
patients were asked via the PGTQ, which uses a 7-point Likert scale ranging from 1 (fully disagree) to 7 (fully agree), about how motivating, frustrating, challenging, and enjoyable they found the training with Negami to be. Since each of the 4 questions covers a different aspect of the perception of Negami, no mean value was calculated.

The patients with spatial neglect received a 2-week application with 5 training sessions per week. Each session lasted 20 to 25 minutes. As the healthy participants, they completed Task A in the first half and Task B in the second half. The tasks were always initiated for the patients with the easiest difficulty level. The criterion for success in advancing to the next difficulty level was that the task was successfully completed 3 times in a row. A return to a lower difficulty level was indicated when the task was not solved 2 times in a row. Upon completion of the treatment, subjective reports about the use of the Negami app were collected from the patients by using the following three questions: (1) How satisfied were you with the treatment? (2) How motivating did you find the treatment? (3) How much did you enjoy the treatment? The answer options varied from 1 to 6, with 1 being the best result. The patients were further asked whether they would recommend the treatment to others (yes/no) and in an open response format, whether they had any suggestions for improvement.

Results

Healthy Participants

The results of the healthy elderly participants from the SUS on usability resulted in a mean value of 4.3 (SD 0.6), indicating that the usability of the Negami app was rated on average as high to very high. Moreover, there were hardly any side effects detected, as assessed by the SSQ and indicated by a mean score of 0.8 (SD 1.32). More specifically, only 1 patient (10%) reported mild symptoms of fatigue, headache, difficulty focusing, and blurred vision. In addition, 2 (20%) other participants reported mild problems with strained eyes and focusing. None reported moderate or strong side effects. Finally, the PGTQ captured the dimensions of motivation, frustration, challenge, and entertainment during training with Negami. The healthy participants’ evaluations on these dimensions are shown in Figure 5. The participants found Negami to be motivating (mean 6.1, SD 0.7) and highly entertaining (mean 6.5, SD 0.9). As indicated by a mean score of 1.5 (SD 0.7), the training was not perceived as frustrating. However, some participants found the training challenging, while others did not (mean 4.1, SD 1.7).

Figure 5. Percentage of healthy, older adults (n=10) evaluating training with the Negami app on the dimensions “motivation,” “frustration,” “challenge,” and “entertainment,” on a 7-point scale (1 = “I fully disagree”; 7 = “I fully agree”).
**Neglect Patients**

An analysis of the reports by the patients regarding the use of the app revealed that all 10 patients recommended treatment with Negami. The patients’ evaluation of the app treatment on the dimensions of fun, satisfaction, and motivation are shown in Figure 6. The results demonstrate that all patients evaluated the app treatment on all 3 dimensions with the best 2 scores. Only 1 (10%) of the patients chose a score of 4 on the dimension of satisfaction. Consistent with these very positive ratings by the patients with spatial neglect, the Spearman rank correlation coefficients between the dimensions of satisfaction, motivation, and fun and the severity of spatial neglect (expressed as the mean value of the CoC from the 2 cancellation tasks) were not significant (satisfaction: r=.21, motivation: r=.57, fun: r=.42; all P≥.086). Regarding suggestions for improvement in the open response format, 3 (30%) of the 10 patients with hemiparetic neglect reported that they had trouble holding the tablet due to the weight. They had to hold the tablet with 1 hand instead of both. One (10%) patient noted that older patients might be digitally overwhelmed with the treatment. All other patients had no suggestions for improvement.

**Figure 6.** Percentage of stroke patients (n=10) with spatial neglect evaluating training with the Negami app on the dimensions “satisfaction,” “motivation,” and “fun,” on a 6-point scale (1 is the best score; 6 is the lowest score).

---

**Discussion**

**Principal Findings**

This study aimed to develop and assess the usability, side effects, and game experience of a new rehabilitation app (Negami) for patients with spatial neglect. By using AR, the app was designed to playfully encourage patients to actively explore space by turning their gaze, head, and trunk. Our investigation showed that Negami was perceived as motivating and entertaining by the healthy elderly controls. Training at the highest defined difficulty level was perceived as differently challenging but not as frustrating. The survey further revealed high usability and no to very low side effects such as headache, strained eyes, or blurred vision. Additionally, the second group evaluated in this study (ie, the patients experiencing spatial neglect after a stroke) consistently evaluated the app positively on the dimensions of motivation, satisfaction, and fun. Furthermore, we observed that the app and the tablet could also be used by patients with hemiparesis (by holding the tablet with the unimpaired hand only, grasping the tablet in the middle from the top or bottom). However, 3 (30%) of the 10 patients with hemiparesis complained about the heaviness of the tablet, so the use of even lighter devices seems beneficial.

**Gamification and the Negami App**

These positive results give us reason to believe that the new app could become a very useful tool in the treatment of patients with spatial neglect. Ogourtsova and colleagues [38] have shown that patient expectation of performance is the most important factor for acceptance of the use of VR technologies in rehabilitation. Therefore, when defining the 3 different difficulty levels of training with Negami, our focus was on ensuring that the easiest task could be solved relatively quickly even for patients with more severe neurological impairment. This could also be the reason why Negami was so well received by the tested patients who had a stroke. The use of different difficulty levels, as well as auditory and visual feedback on performance during training, are elements that supported the gamification of the Negami app. Gamification for the purpose of rehabilitation has been shown to have positive effects on the patients’ engagement and motivation for their treatment [39]. Another
reason for the positive feedback from both patients and healthy participants could be due to our app design in that we paid special attention to the age group, as 75% of patients who have had a stroke are aged over 65 years [40] and are typically not ‘digital natives’ [41]. In implementing short distances in operation, simple instructions, and assistance, we aimed to ensure that this age group feels comfortable in dealing with an electronic device such as a tablet when using the Negami app.

AR Versus VR Techniques and Telerehabilitation

A third aspect that might have contributed to the positive ratings by the 2 groups tested could be the choice of the AR over VR technique. It is a frequent observation that the use of VR may evoke symptoms that are summarized under the term ‘cybersickness’ [42]. These include symptoms such as drowsiness, headache, balance and coordination problems, nausea, and blurred vision. We observed that such symptoms were absent during the training with Negami. The reason for this could be that the risk of sensory mismatch concerning visual, vestibular, and proprioceptive information, which is considered one of the causes of cybersickness [43], is reduced to a maximum—if not completely avoided—when working with AR. The AR-based Negami app prompts people to move in the real world by actually moving their heads and bodies. The use of the tablet merely adds a virtual object to the real world. In this respect, manipulation is kept to a minimum.

Negami can also offer an advantage over VR games in terms of cost. The development of a VR system is often associated with high costs due to the need for dedicated hardware [42]. Since Negami is available online (see the Negami website), it can be easily downloaded to any available tablet or smartphone. This provides wide availability and ensures that additional costs are avoided. The use of already existing technical hardware also offers immense opportunities regarding remote/telerehabilitation. This is important because, in one-third of patients experiencing spatial neglect in the acute phase of the stroke, the disorder becomes chronic [44]. During their inpatient stay, patients are well cared for in terms of their recovery, but what happens after discharge from inpatient rehabilitation? An answer to that could be remote/telerehabilitation. Easy-to-understand instructions for the tasks in Negami can allow patients to practice independently from home. With online data storage, the therapist has external access to patients’ treatment data and can monitor them remotely.

Limitations

One limitation of this study is the small sample size. We only could include 10 elderly neurologically healthy individuals and 10 patients experiencing spatial neglect and left-sided hemiparesis after a stroke in the right brain hemisphere. Therefore, testing the Negami app with a larger number of participants in a future study will be useful. It would also be desirable to evaluate the use of the app on the part of the therapists in the future. It should be further mentioned that Negami thus far can only be used on Apple devices. However, it is being further developed with the aim of being able to use on Android devices as well. The further development of all smart devices with different operating systems would also ensure the use of lighter devices, as 3 (30%) of the patients complained about the weight of the tablet. At this point, it should also be mentioned that very severely affected patients, who are not yet able to hold a tablet in the acute phase after a stroke, may not benefit from Negami initially, and other therapeutic approaches should be considered first for them.

Conclusion

AR represents a promising extension to the traditional ‘visual exploration training’ for spatial neglect. The natural interaction with the environment during playful tasks appears to markedly increase patient motivation. An advantage of AR, as implemented in Negami, over immersive VR-based treatments may be that it minimizes the risk of sensory mismatch concerning visual, vestibular, and proprioceptive information, which is considered one of the causes of cybersickness.

Acknowledgments

This work was supported by Hector Stiftungen, Weinheim (M2107).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic and clinical data of all 10 patients with right brain damage and spatial neglect.

[DOCX File , 19 KB - games_v11i1e40651_app1.docx ]

Multimedia Appendix 2

Patient performing Task B on Negami.

[MP4 File (MP4 Video), 21397 KB - games_v11i1e40651_app2.mp4 ]

References


32. What is Xamarin? Microsoft. URL: https://docs.microsoft.com/en-us/xamarin/get-started/what-is-xamarin [accessed 2023-02-02]

33. iOS app architecture. Microsoft. URL: https://learn.microsoft.com/en-us/xamarin/ios/internals/architecture [accessed 2023-02-02]


Abbreviations

- API: application programming interface
- AR: augmented reality
- ARM: Advanced Reduced Instruction Set Computer Machine
- CoC: Center of Cancellation PGQT: Perception of Game Training Questionnaire
- EWB: end point weightings bias
- IL: Intermediate Language
- SQL: Structured Query Language
- SSQ: Simulator Sickness Questionnaire
- SUS: System Usability Scale
- UI: user interface
- VR: virtual reality
- XAML: Extensible Application Markup Language

https://games.jmir.org/2023/1/e40651 JMIR Serious Games 2023 | vol. 11 | e40651 | p.409 (page number not for citation purposes)
Stammler B, Flammer K, Schuster T, Lambert M, Karnath HO
Negami: An Augmented Reality App for the Treatment of Spatial Neglect After Stroke
JMIR Serious Games 2023;11:e40651
URL: https://games.jmir.org/2023/1/e40651
doi:10.2196/40651
PMID:36848215

©Britta Stammler, Kathrin Flammer, Thomas Schuster, Marian Lambert, Hans-Otto Karnath. Originally published in JMIR Serious Games (https://games.jmir.org), 27.02.2023. 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.
Original Paper

Appropriate Image Selection With Virtual Reality in Vestibular Rehabilitation: Cross-sectional Study

Kerem Ersin¹*, MSc; Emre Gürlek¹*, BSc; Hakan Güler¹*, BSc; Çiğdem Kalaycık Ertugay²*, Assoc Dr; Mustafa Bülent Şerbetçioglu³*, Prof Dr

¹Graduate School of Health Sciences, Istanbul Medipol University, Istanbul, Turkey
²Department of Audiology, Istanbul Education Research Hospital, Istanbul, Turkey
³Faculty of Health Sciences, Istanbul Medipol University, Istanbul, Turkey

*all authors contributed equally

Corresponding Author:
Emre Gürlek, BSc
Graduate School of Health Sciences
Istanbul Medipol University
Göztepe Mahallesi Atatürk Caddesi No. 40 Beykoz
Istanbul, 34820
Turkey
Phone: 90 507 934 71 16
Fax: 90 212 521 23
Email: emregurlek1622@gmail.com

Abstract

Background: While vestibular rehabilitation with virtual reality (VR) is becoming more popular every day, the disadvantages of this method are not yet clear.

Objective: The aim of this study is to examine the effect of the image to be used in vestibular rehabilitation with VR on the systems that provide body posture.

Methods: The study was carried out with 36 participants (18 women and 18 men) aged 18 to 30 years. To assess balance control components separately, a sensory organization test was administered to the participants in the presence of stressful and relaxing environment images with VR technology. The State-Trait Anxiety Inventory survey was also used to measure the stress values in the created environments.

Results: The State-Trait Anxiety Inventory survey revealed that while stressful videos significantly increased stress, relaxing videos reduced stress. Among measurements obtained in the presence of VR, significant decreases were observed mostly in the visual system data. A significant increase in vestibular system data ($P=0.01$) was observed with a decrease in visual system data ($P<0.001$) when the relaxing image was presented. Additionally, there was a significant difference in the somatosensory ($P=0.001$), composite ($P=0.002$), and visual system ($P<0.001$) data in the presence of stressful videos.

Conclusions: Although the use of a VR system for vestibular rehabilitation is relatively new, no extant studies have examined how the image type used in VR can affect the integration of visual system data. Therefore, this study is unique in terms of showing the effects of the stress created by the change in the type of the image used in VR. When VR technology is used for therapeutic vestibular rehabilitation for patients whose balance disorder is due to the vestibular system, stress-free videos should be used. However, the use of stressful videos in VR technology will be beneficial in the rehabilitation of those with balance disorders due to the somatosensory system.

(JMIR Serious Games 2023;11:e40806) doi:10.2196/40806

KEYWORDS

balance; computerized dynamic posturography; stress; vestibular rehabilitation; vestibular system; virtual reality


### Introduction

Maintaining posture uses a complex mechanism involving the brain’s processing of multiple sensory inputs from the visual, proprioceptive, and vestibular systems by the central nervous system [1]. Computed dynamic posturography (CDP), which is used to examine the effects of these multisensor inputs separately, is a kind of test that can be used to evaluate the mechanisms that ensure the protection of the individual’s posture by using different test positions similar to those we often experience in our daily lives [2]. The sensory organization test (SOT), one of the CDP test protocols, is a test that can evaluate a person’s ability to use inputs from the vestibular, somatosensory, and visual systems to maintain balance in different stimulus situations and to monitor the compensatory condition in case of balance deterioration [3].

Virtual reality (VR) can be defined as a computer-generated simulation of a 3D image or environment with which the user can interact. Recently, the scope of VR use has expanded greatly, and it is being used in many areas, especially in rehabilitation programs, where it is increasing the effectiveness of vestibular rehabilitation. Studies have shown that a VR protocol may be a safe option to improve postural control and quality of life in individuals with vestibular disorders [4-6].

The classical stress response is a negative feedback system mediated by the hypothalamic-pituitary-adrenaline axis [7]. Studies have found that balance is affected in people who are under stress [8-10]. According to Saman et al [11], stress may affect central vestibular function either directly through the effects of cortisol and corticosteroids on ion channels and neurotransmission in the brain or indirectly through stress-related neuroactive substances. Gliddon et al [12] observed a significant increase in postoperative salivary cortisol levels in guinea pigs undergoing unilateral vestibular deafferentation.

One of the most widely used clinical tests for the measurement of anxiety is the State-Trait Anxiety Inventory (STAI) survey (Multimedia Appendices 1 and 2). The survey test, which evaluates stress immediately after an individual has undergone a test in a research or clinical setting, was administered, and the Turkish validity form, the reliability of which had been evaluated [16]. In the first condition, the eyes are open and the visual environment moves to present incorrect visual sequences. In the fourth to sixth conditions, the force plates are moved so that they act as what is referred to as a “sway referenced surface,” making proprioceptive information inaccurate; hence, the subjects use visual and vestibular (or only vestibular) information to maintain their balance, in addition to having to overcome the inaccurate proprioception signals by relying on other systems and compensatory mechanisms. In the fourth condition, information from the visual and vestibular systems is evaluated. In the fifth condition, the eyes are closed.

### Methods

#### Participants and Recruitment

We calculated the sample size using power analysis. In the study, we set an α (type I error) of 5%, Cohen medium standardized effect size of 0.5, and power of 80% (with β [type II error] set to 20%). Power analysis revealed that a minimum of 28 patients were required. A total of 36 patients, 18 men (aged 23.33, SD 2.54 years) and 18 women (aged 21.66, SD 2.22 years) between the ages of 18 and 30 years (sample mean age 22.50, SD 2.50 years), were included in the final analysis. While individuals who were psychologically healthy and had no problems with balance in daily life in the preceding 1-year period were included in the study, those with eye problems, a fear of heights, dizziness in daily life, physical discomfort, or regular use of alcohol or drugs were excluded from the study.

#### Experimental Design

The study was conducted in the audiology clinic at the Istanbul Training and Research Hospital between February 23 and May 10, 2021. A NeuroCom device (Natus Medical Inc) was used to administer the SOT from among the CDP test repertoire. The SOT measures how well the patient maintains equilibrium under 6 sensory conditions. Each condition is scored between 1 and 100, with an increased score indicating better stability. Each of the 6 conditions is repeated 3 times, and the duration of each test is 20 seconds. After 6 conditions (ie, 18 tests) the composite equilibrium score is then determined [16].

For the first 3 conditions, the force plates are fixed, and for the other 3 conditions, they move in anterior and posterior directions. In the first condition (baseline), the person is placed on the system such that all sensory information involved in postural control is available. In the second condition, the participant is tested with eyes closed (the visual information system is eliminated). In the third condition, the person’s eyes are open, but the visual environment moves to present incorrect visual sequences. In the fourth to sixth conditions, the force plates are moved so that they act as what is referred to as a “sway referenced surface,” making proprioceptive information inaccurate; hence, the subjects use visual and vestibular (or only vestibular) information to maintain their balance, in addition to having to overcome the inaccurate proprioception signals by relying on other systems and compensatory mechanisms. In the fourth condition, information from the visual and vestibular systems is evaluated. In the fifth condition, the eyes are closed.

In this condition, information from the vestibular system is evaluated. In the sixth condition, inappropriate visual arrays are presented; thus, information from the vestibular system is evaluated [16].

Comparison of scores from the first 2 trials produces a “somatosensory score”; comparison of the score for test 4 with that from test 1 yields the “visual score”; comparison of scores from trial 5 to trial 1 reflects the “vestibular score.” Comparing the sum of scores from the third and the sixth trials with the sum of scores from the second and fifth trials yields the “visual preference score” [17].
The SOT, one of the CDP tests, was administered to the participants in 3 different conditions generated using VR: in the control condition, in a stressful environment, and in a relaxing environment (Figure 1). An HTC VIVE VR system was used to generate the different environments. In the control condition, normalization values were determined by administering the SOT to the participants without wearing VR glasses (Figure 1).

To generate a stressful environment in the VR simulation, participants were raised to a height of 40 meters in a simulation of an elevator. Then, in the simulation, the elevator door opened, and they walked onto a platform that appeared in front of them. The end of the platform was mapped to the test area of the CDP device. When they arrived at this position, the SOT was administered to the participants while they saw VR images from a simulated height of 40 meters (Figure 1).

In the relaxing environment, while participants were waiting for the test on the CDP device, a meditation application was opened in VR. The SOT was administered while the participants were in the relaxing scenery (Figure 1).

Considering that the participants could adapt to the test, the tests were applied to each participant starting from a different condition. This aimed at preventing possible adaptation. A participant who completed one condition rested for at least 1 hour, after which another condition was applied.

After the tests, the participants took the STAI survey. Certain questions in the survey could increase or decrease the anxiety score. For each question, a score between 1 (or –1) and 4 (or –4) is assigned in accordance with the positive or negative characteristics of the question. The highest score is 80 and the lowest score is 20. The higher the total anxiety score, the higher the anxiety level of the person taking the survey [13].

Figure 1. Three different visually generated test environments.

Ethics Approval
The study was approved by the Non-Invasive Clinical Research Ethics Committee of Istanbul Medipol University (923/2020). All individuals participating in the study received explanations of the aims of the study, how long it would last, and practices and expectations. Participants then signed the consent form.

Statistical Analysis
The minimum subject size was estimated using G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf). The data analysis of our study was performed using SPSS (version 22.0; SPSS Inc). Descriptive statistics included mean and SD values. Normality of the distribution and homogeneity of the data were analyzed with the Kolmogorov-Smirnov test. A triple comparison of the groups (control condition [CC], stressful condition [SC], and relaxing condition [RC]) was carried out using the Friedman test. CC-RC, CC-SC, and RC-SC comparisons were made using the Wilcoxon signed-rank test. The statistical significance level was set to a P value of ≤ .05 when analyzing the results of the Wilcoxon signed-rank test. In the Friedman test, statistical significance level was set to a P value of ≤ .017.

Results
Triple Comparisons
Significant differences were observed in the visual system (P < .001) and somatosensory system (P = .001) data in the triple comparison of SOT data obtained in different stress environments (Figure 2). While no significant difference was observed in the vestibular system (P = .03) data, a significant difference was observed in the composite (P = .002) values (Figure 3). While a significant difference was observed in stress values (P < .001), there was no significant difference in the preference values (Figure 4).
Figure 2. The sensory organization test and visual and somatosensory system data.

Figure 3. The sensory organization test for the vestibular system and composite data.

Figure 4. The sensory organization test, preference scores, and survey stress data.

**Dual Comparisons**

Significant differences were observed in visual system data in pairwise comparisons between the cases (CC-RC and CC-SC, $P<.001$; RC-SC, $P=.001$). A significant difference was observed when comparing the CC and SC environments ($P=.002$) in the somatosensory system data. A significant difference was observed when comparing the CC and RC environments ($P=.002$) in the vestibular system data (Table 1). No significant difference was observed in the comparisons made with preference data. A significant difference was observed when comparing the CC and SC environments ($P=.02$) and the SC and RC environments ($P=.005$) in the composite data. When only the stress data were evaluated, a significant difference was observed when comparing the CC and SC environments ($P<.001$), CC and RC environments ($P<.001$), and SC and RC environments ($P<.001$; Table 2).
when we looked at the instantaneous effect of VR on the systems that maintain posture by comparing the CC and RC environments and considering the results of the SOT, we found that VR application does not directly cause stress, but rather makes the person feel more comfortable than normal. However, we also observed that VR had a negative effect on the visual system while in this condition. Although there was no negative effect on the somatosensory system, we observed a significant increase in the values of the vestibular system. With the decrease in values for the visual system and the increase in those for the vestibular system, there was no significant decrease in the composite value, which is the general balance score. We believe that VR-based vestibular rehabilitation will be a successful process based on these findings. To avoid potential side effects from staying in a virtual environment for a long time during this process, special attention should be paid to the effect on the visual system, cyber sickness, dizziness under visual stimuli, nausea, and imbalance [24,25].

In Öztürk’s [25] study, individuals were evaluated with cervical vestibular-evoked myogenic potentials and video head impulse tests in the presence of visual illusions. The amplitudes of the vestibular-evoked myogenic potentials increased in cases where visual illusions were presented, and gains in some semicircular canals increased significantly. This showed that the vestibular system works more efficiently when there is a decrease in the consistency of the data obtained from the visual system [25]. In this study, the increase in vestibular system data is compatible with the visual stimulus provided by wearing the VR device.

In a study by Smyth et al [26], healthy individuals with the highest cortisol response to psychosocial stress—that is, those who were most affected—had less-impaired postural sway when exposed to visual movement [26]. This is consistent with the findings of our study, in which the somatosensory system performed better when the VR image containing the stress stimulus was used, because in healthy people, minor oscillations are modulated by the ankle, which is part of the somatosensory system.

Psychological effects are common in patients experiencing vertigo or dizziness. In particular, the fear of falling and the

### Table 1. Dual and triple comparisons of visual, vestibular, and somatosensory system data in 3 different cases.

<table>
<thead>
<tr>
<th>System</th>
<th>Condition</th>
<th>Mean (SD)</th>
<th>P value (SC-RC)</th>
<th>P value (CC-SC)</th>
<th>P value (CC-RC)</th>
<th>P value (SC-RC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual system</td>
<td>Control</td>
<td>86.88 (8.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>65.22 (19.98)</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>75.30 (13.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vestibular system</td>
<td>Control</td>
<td>68.36 (12.20)</td>
<td>.03</td>
<td>.43</td>
<td>.01</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>70.11 (16.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>73.97 (13.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatosensory system</td>
<td>Control</td>
<td>97.22 (8.79)</td>
<td>.001</td>
<td>.002</td>
<td>.21</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>100.16 (5.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>99.11 (3.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Dual and triple comparisons of preference, composite, and stress data in 3 different cases.

<table>
<thead>
<tr>
<th>Session</th>
<th>Condition</th>
<th>Mean (SD)</th>
<th>P value (SC-RC)</th>
<th>P value (CC-SC)</th>
<th>P value (CC-RC)</th>
<th>P value (SC-RC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference</td>
<td>Control</td>
<td>99.19 (9.95)</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>102.25 (8.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>101.55 (6.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composite</td>
<td>Control</td>
<td>76.91 (7.62)</td>
<td>.002</td>
<td>.02</td>
<td>.52</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>74.11 (9.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>77.44 (8.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>Control</td>
<td>33.50 (8.48)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Stressed</td>
<td>40.16 (12.18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relaxed</td>
<td>26.38 (6.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

**Principal Findings**

The aim of the study was to investigate the effect of stress on the systems that provide data for maintaining posture. The SOT was administered to the study participants in different stress environments generated with VR. The changes in visual, vestibular, and somatosensory systems were examined with the SOT. Our results indicate that stress, especially presented visually, has a significant effect on the systems that protect posture.

Wuehr et al [18] took people to a height of 40 meters with a VR-based elevator simulation and observed specific changes in subjective fear levels, muscle activity, and balance control. In our study, like that of Wuehr et al [18], the SOT was administered by raising people to a height of 40 meters in an elevator simulation with a VR system as an instant stress source. When we examined the change in stress in the participants using the STAI survey before and after the test, we found that the participants were affected by stress in the presence of a stressful video. In the presence of a relaxing video, we found that the stress level was significantly lower than in the control group. Thus, we proved that our image selection is effective for assessing stress factors.

VR, which we use as an instant stress generation tool, has been used frequently in clinical studies with the development of technology in recent years [19]. In addition, its use in vestibular rehabilitation is increasing [20]. It has not been specifically reported in the literature whether VR has an adverse effect in VR-based vestibular rehabilitation or whether it has an adverse effect on the systems that provide data for maintaining posture [21]. Micarelli et al [22,23] reported that symptoms of nausea, oculomotor stress, and disorientation decreased over time in people participating in a vestibular rehabilitation program with VR. Both studies saw a significant reduction in nausea, oculomotor stress, and disorientation in the first to fourth week of the rehabilitation program with VR [22,23]. However, the instantaneous effect of VR and the effect of the video type used on treatment were not investigated in these studies. In our study,
resulting stress and anxiety seriously impair the quality of life of these individuals [27]. Egger et al [28] found that patients with vestibular symptoms had higher social stress and trait anxiety scores than those without vestibular symptoms. The balance-stress neuroanatomical link has been defined as follows: the nucleus tractus solitarius has been shown to have extensive connections with the vestibular nuclei, both directly and through indirect projections through the parabrachial nucleus, which provides major inputs to the limbic system, including the enlarged central amygdaloid nucleus, infralimbic cortex, and hypothalamus [29]. Because of this connection between the vestibular system and the limbic system, the stress factor should be taken into account when planning treatment for individuals with vestibular symptoms. Choosing less stressful images during the vestibular rehabilitation phase, especially near the end of the study, will aid in the recovery of vestibular systems.

Studies have shown that stress has an effect on the systems that maintain posture [7,9,11]. In line with the findings of the literature review, this study found that stress has a significant effect on these systems and that the stress effect provided by VR is primarily due to the effect on the visual system. When comparing the RC and SC environments, there was a significant decrease in visual system data and composite scores, accompanied by a significant increase in stress values. When we compared the visual system values of the CC and SC environments, we found a significant decrease as a result of stress, as well as a significant change in stress. Similarly, a significant increase in somatosensory values was observed with stress. This result shows the importance of the image used in vestibular rehabilitation. Confidence in the somatosensory system increased significantly when the participants watched a stressful video, while confidence in the vestibular system increased significantly when a relaxing video was shown. Thus, data on which group would be more effective for use in vestibular rehabilitation with VR, and the appropriate images to be used, should be highlighted.

Limitations
The SOT is not normally administered while wearing VR glasses. However, to investigate the effects of VR on balance while wearing these glasses, the SOT was used with relaxing images. To investigate the effect of image selection, a stress-inducing image, rather than a relaxing image, was used.

Conclusions
Our study investigated the effects of stress on the systems that protect posture and how much each of the systems is affected by stress, and the effect on individuals of using VR technology to provoke stress. Our results show that the visually created stressful environment had a significant effect on balance. In individuals with balance disorder, it would be more beneficial to use stressful images if the problem relates to the somatosensory system and to use less stressful images if it is related to the vestibular system. To obtain more successful and quicker results in the treatment processes for individuals with balance problems, it is necessary to focus on the stress condition and its effect on patients. In addition, the decrease in visual system data as a result of the stress we induced visually and the effect of rehabilitation with VR in people whose visual system contributes to their balance disorder should be the subject of further studies.

Acknowledgments
We would like to thank the participants for their patience and invaluable contribution to this study. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Stait-Trait Anxiety Inventory.
[PDF File (Adobe PDF File), 43 KB - games_v11i1e40806_app1.pdf ]

Multimedia Appendix 2
Stait-Trait Anxiety Inventory (Turkish).
[PDF File (Adobe PDF File), 71 KB - games_v11i1e40806_app2.pdf ]

References


14. Spielberger CD, Gonzalez-Reigosa F, Martinez-Urrutia A, Natalicio LFS, Natalicio DS. The State-Trait Anxiety Inventory. Interim J Psychol 1971;5(3&4) [FREE Full text]


Abbreviations

CC: control condition
CDP: computed dynamic posturography
RC: relaxing condition
SC: stressful condition
SOT: sensory organization test
STAI: State-Trait Anxiety Inventory
VR: virtual reality
Cognitive Training With Head-Mounted Display Virtual Reality in Neurorehabilitation: Pilot Randomized Controlled Trial

Julian Specht1*, BSc; Barbara Stegmann1*, BSc; Hanna Gross2*, BSc; Karsten Krakow2,3*, Prof Dr

1SRH University of Applied Sciences Heidelberg, Department of Applied Psychology, Heidelberg, Germany
2Asklepios Neurologische Klinik Falkenstein, Department of Neurorehabilitation, Königstein im Taunus, Germany
3Rehaklinik Zihlschlacht, Department of Neurorehabilitation, Zihlschlacht, Switzerland
*all authors contributed equally

Corresponding Author:
Julian Specht, BSc
SRH University of Applied Sciences Heidelberg
Department of Applied Psychology
Ludwig-Guttmann-Straße 6
Heidelberg, 69123
Germany
Phone: 49 01737808422
Email: julian.specht@livingbrain.de

Abstract

Background: Neurological rehabilitation is technologically evolving rapidly, resulting in new treatments for patients. Stroke, one of the most prevalent conditions in neurorehabilitation, has been a particular focus in recent years. However, patients often need help with physical and cognitive constraints, whereby the cognitive domain in neurorehabilitation does not technologically exploit existing potential. Usually, cognitive rehabilitation is performed with pen and paper or on a computer, which leads to limitations in preparation for activities of daily living. Technologies such as virtual reality (VR) can bridge this gap.

Objective: This pilot study investigated the use of immersive VR in cognitive rehabilitation for patients undergoing inpatient neurorehabilitation. The goal was to determine the difference in rehabilitation effectiveness between a VR serious game that combines everyday activities with cognitive paradigms and conventional computerized cognitive training. We hypothesized the superiority of the VR serious game regarding cognitive abilities and patient-reported outcomes as well as transfer to daily life.

Methods: We recruited 42 patients with acute brain affection from a German neurorehabilitation clinic in inpatient care with a Mini Mental Status Test score >20 to participate in this randomized controlled trial. Participants were randomly assigned to 2 groups, with 1 receiving the experimental VR treatment (n=21). VR training consisted of daily life scenarios, for example, in a kitchen, focusing on treating executive functions such as planning and problem-solving. The control group (n=21) received conventional computerized cognitive training. Each participant received a minimum of 18 treatment sessions in their respective group. Patients were tested for cognitive status, subjective health, and quality of life before and after the intervention (Alters-Konzentrations-Test, Wechsler Memory Scale—Revised, Trail Making Test A and B, Tower of London—German version, Short Form 36, European Quality of Life 5 Dimensions visual analog scale, and Fragebogen zur Erfassung der Performance in VR).

Results: Repeated-measures ANOVA revealed several significant main effects in the cognitive tests: Tower of London—German version (P=.046), Trail Making Test A (P=.01), and Wechsler Memory Scale—Revised (P=.006). However, post hoc tests revealed that the VR group showed significant improvement in the planning, executive control, and problem-solving domains (P=.046, Bonferroni P=.02). In contrast, no significant improvement in the control group between t0 and t1 was detected (all P>.05). Furthermore, a nonsignificant trend was observed in visual speed in the VR group (P=.09, Bonferroni P=.02).

Conclusions: The results of this pilot randomized controlled trial showed that immersive VR training in cognitive rehabilitation had greater effectiveness than the standard of care in treating patients experiencing stroke in some cognitive domains. These findings support the further use and study of VR training incorporating activities of daily living in other neurological disorders involving cognitive dysfunction.

Trial Registration: Federal Registry of Clinical Trials of Germany (DRKS) DRKS00023605; https://drks.de/search/de/trial/DRKS00023605
Introduction

Background

The primary goal of neurorehabilitation is to restore or preserve critical function, thereby enabling patients to participate in daily life. Neurological diseases can lead to severe disabilities. This requires the patient to make great effort in rehabilitation. For example, stroke is one of the leading causes of disability. It is estimated that 33% to 42% of stroke survivors require assistance in managing activities of daily living (ADL). Reduced ability to manage ADL is often long term, with 20% of patients being unable to manage ADL independently 5 years after a stroke [1,2]. Even patients with only mild stroke have unmet rehabilitation needs and require support in managing ADL [3,4]. This also raises the question of whether there should be more ADL-focused therapies in principle, because the study by Rejnö et al [1] showed that there is a continuous decrease in ADL independence over time after a stroke; the proportion of patients who can independently manage ADL 3 months after a stroke decreases by about 3% per year, so that 5 years after the stroke, only just over 80% of the group of independent patients can independently manage ADL [1].

Furthermore, stroke is associated with a substantial economic burden, creating a cost of €60 billion (US $67.2 billion) in the European Union [5]. Wilkins et al [6] highlighted that these costs are far greater when including indirect costs from the subsequent burden on society by lost employment and expenses related to the need for multidisciplinary, extensive rehabilitation. This points to the great need for more effective rehabilitation interventions that support patients in regaining their ability to participate and manage everyday life independently.

ADL, such as making coffee or preparing lunch, require multiple motor and cognitive processes to achieve their functional goals. Seemingly trivial activities demand several temporally overlapping cognitive functions such as planning, problem-solving, recall, and execution [7]. Toth et al [8] demonstrated in collaboration with 36 occupational therapists that the 20 most critical instrumental ADL were closely related to different cognitive functions in older adults. Cognitive domains that have been frequently discovered in this context are memory, executive functions (EF), attention, visuospatial abilities, and language. They also found that, for instance, in grocery shopping, making and keeping medical appointments and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanded cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demanding cognitive domains, whereas paying bills, making a phone call, and medication management EF were the most demands cognitive domains, whereas paying bills, making a phone call, and making and keeping medical appointments were identified to require language the most [8]. In addition, there exists a correlation between ADL performance, multimorbidity, and quality of life [9]. Therefore, it would be beneficial if neurorehabilitation methods emphasized preparation for everyday life to reduce morbidity. One component of this process is improving the cognitive skills required to handle ADL. With the findings of Toth et al [8], it would be beneficial if rehabilitation explicitly addresses patients’ requirements in everyday life and provides training, especially on the more complex instrumental ADL and its cognitive functional domains.

However, computerized cognitive training (CCT), or pen-and-paper training, is the standard practice in German rehabilitation facilities. It involves training in isolated domains, including attention, working memory, EF, and language. They are economical and efficient tools for delivering therapies. Still, its use is recommended only with restrictions by national guidelines for treating EF in various diseases, for example, in the stroke guidelines [10].

Furthermore, researchers discovered insufficient evidence of computer-based training in a second-order meta-analysis in neurorehabilitation [11]. The authors concluded that the processes initiated by CCT are often limited to near-transfer effects. The effects remained within the scope of the trained task regardless of the far-transfer measure, study population, or cognitive training application used. However, a far-transfer effect is crucial for later-stage rehabilitation and return to daily life. This report aligns with the above-referenced stroke guidelines, advising caution using computer-based training owing to the limited evidence on its effectiveness. Although Gajewski et al [12] demonstrated in their clinical trial that some far-transfer effect occurs when conducting combined PC and paper-and-pencil training, a transfer effect on ADL was not detectable.

Instead, customized virtual reality (VR) training is suitable for multidomain ADL training [13,14]. VR is defined as images and sounds created by a computer that seem almost indistinguishable from the real world and users can interact with via sensors [15]. Although it can be experienced nowadays with purpose-made head-mounted displays (HMDs) placed on the head in front of the eyes for a fully immersive experience, the interaction with the virtual world is mainly carried out with 2 hand controllers. VR enables a combination of principles from occupational therapy, neuropsychology, and ADL, resulting in an ecologically valid training setting for many individuals with neurological impairments who want to regain their ability to participate in daily activities. Although insights exist concerning the acceptance of immersive HMD-VR in patients with stroke [16] or older adults [17], knowledge about the processes of cognitive rehabilitation facilitated by immersive VR needs to be expanded.

One concept that is becoming more widespread in neurorehabilitation is serious games (SG) [18]. SGs are defined as games developed not only for entertainment but also for serious learning purposes and can be considered a nonpharmacological, gamified approach to treat and evaluate [19] a patient’s executive dysfunction in an ecologically valid way [20]. SGs and VR have already been combined, as Faria et al [21] demonstrated. They trained EF in immersive VR with a virtual city, Reh@City, including attention, memory,
visuospatial abilities and other EF, in performing ADL and found that this approach boosts cognitive rehabilitation and leads to better results in contrast to conventional cognitive rehabilitation methods [21]. A systematic review of the assessment and treatment of EF with VR in clinical populations and healthy participants concluded that VR is a promising tool [22]; however, more research is necessary, as the authors cited only a few trials on this topic. Training systems focused on ADL have the potential for better transfer to the real world after rehabilitation. Still, more evidence on the interrelation between SGs, ADL, EF training, and immersive VR [22] is necessary.

Investigating VR-based cognitive rehabilitation involves research of a multitude of groups. The number of publications in scientific journals has increased almost exponentially since 2004 [23]. The 3 most frequently studied diseases in VR-based cognitive rehabilitation are stroke, dementia, and mild cognitive impairment. Jahn et al [13] noted in their systematic review of publications about VR and cognitive rehabilitation that there is promising evidence for the functionality of this approach. However, current research findings still struggle with several issues, including group sizes being too small (N=6-34 participants, with 89% of reviewed trials having N<20 participants per intervention arm), the lack of randomization (eg, no randomization at all), or suboptimal statistical analyses (no group-by-time interaction analysis reported).

Maggio et al [14] found in their review of VR for cognitive rehabilitation that VR training in neurorehabilitation has measurable effects on executive functioning, visuospatial abilities, speech, attention, and memory skills. In particular, executive functioning and depression at discharge from stroke rehabilitation are strong predictors of functioning at the time of discharge and the following 12 months, as Shea-Shumsky et al [24] found in their study that executive functioning was a strong predictor of ADL capacity to a more significant extent than mental status. It is a strong signal that problems with EF are involved in issues regarding functionality in everyday life and that viable training is helpful and necessary [24].

**Objectives**

In this paper, we present the results of a pilot clinical randomized controlled trial in patients undergoing neurological treatment after acute brain injury. The trial compared two treatments: (1) a novel immersive VR intervention to treat cognitive dysfunction, specifically EF, presenting ADL in the form of an SG, and (2) conventional computerized cognitive rehabilitation training in inpatient care at a German rehabilitation hospital. VR intervention includes training in a virtual kitchen or garden, simultaneously targeting various cognitive domains, including planning, attention, and problem-solving. Furthermore, interaction with the virtual environment was enabled using HMD-VR to fully immerse users. The trial was conducted to investigate whether training with a VR-SG produces measurable effects in different cognitive domains in patients after acute brain injury and whether the potential effects outperform, are similar to, or are inferior to traditional training methods, specifically CCT.

We hypothesized that VR training will significantly improve cognitive performance and outcomes related to quality of life and general health compared with our control condition throughout the 4-week treatment. Therefore, this study addressed the following objectives:

- The primary objective was to assess the impact of an ADL-based VR-SG on patients’ cognitive abilities after acute brain injury throughout a 4-week treatment period and compare it with a conventional CCT treatment.
- The secondary objective was to assess the impact of an ADL-based VR-SG on self-reported outcomes, quality of life, state of health, and affect and compare it with conventional CCT treatment.
- The tertiary objective was to assess the impact of an ADL-based VR-SG on the actual transfer of learned abilities in VR to daily life and reality compared with conventional CCT treatment.
- The exploratory objective was to investigate the extent to which an ADL-based VR-SG is suitable for cognitive rehabilitation from phase C onward.

No changes to these objectives were made after trial commencement.

**Methods**

**Participants**

Participants were screened and recruited for enrollment at the Asklepios Neurologische Klinik Falkenstein clinic in Königstein im Taunus, Germany. The patients were transferred to the rehabilitation clinic directly from the emergency hospital (rehabilitation phase A) or had already undergone early rehabilitation (rehabilitation phase B). The differentiation to which unit a patient was admitted was based on the Barthel index: it was first developed by Mahoney and Barthel in 1965 [25] and Colin et al [26] and Shah et al [27] modified it over the years. The original 10-item form consists of 10 scales describing ADL, including feeding, bathing, grooming, dressing, bowel and bladder control, toilet use, transfers (bed to chair and back), mobility, and stair climbing. The rating describes the independence of patients to perform the named ADL on a scale from 0 to 10, with a maximum of 15 points for the transfers and mobility scale. The maximum score was calculated by summing up the results of all the scales. A maximum score of 100 indicates that patients have no difficulty managing the described ADL, whereas a score between 30 and 70 indicates the patient’s ability to conduct rehabilitation phase C in Germany. All patients were admitted to the rehabilitation phase C unit and had subacute stroke syndromes. The trial was planned using a parallel design.

The admission interview at the rehabilitation clinic and records of newly admitted patients undergoing rehabilitation were used to recruit the sample. A phase C status qualified patients to participate in the clinical trial, and a transition to phase D during study participation was possible. None of the patients started earlier or later than phase C. All participants were required to be legally competent and of age and had to undergo inpatient neurological rehabilitation treatment at the neurorehabilitation clinic because of acute brain affection (eg, cerebral infarction, intracerebral hemorrhage, subarachnoid hemorrhage, subdural intraparenchymal hemorrhage).
or epidural intracranial hemorrhage, traumatic brain injury, brain tumor, meningitis or encephalitis, and hypoxic encephalopathy), had to have full functionality for at least one arm, had known adequate or corrected eyesight in the central field of vision, and had to give written consent to participate. Patients were also screened for cognitive performance using the Mini Mental Status Test (MMST) [28]. The score obtained had to be >20 out of 30 points. The cut-off value was chosen because some cognitive impairment was expected in this sample, and scores of 27 to 30 (no cognitive impairment) would not properly represent the sample of patients undergoing neurorehabilitation. Patients were excluded if they had present or prior psychiatric comorbidities or disorders and a known overreaction to visual stimulation leading to seizures. Once patients eligible for the study were identified, they were targeted by trained study personnel and informed of the opportunity to participate. In this course, the content and objectives of the study are explained.

Materials: Training Tools

For the trial, 2 different treatment methods were applied. The control group received CCT used in routine care at the rehabilitation center at least 3 times a week for 30 minutes each session, in this case, Freshminder. Freshminder is a cognitive training program with multiple training paradigms, focusing on attention, planning, and memory (Figure 1). Training was conducted on a PC with a keyboard and mouse. Users must interact with the application by clicking objects or deciding by pressing a key on the keyboard. In this study, the following tasks were considered: pearlfish (vigilance and working memory), picture series (memory), step sequence (action planning), task switching (flexibility), and double play (divided attention). For example, in the pearlfish task, participants had to click on fish carrying pearls. Participants had to discriminate which fish were holding the pearls and identify the fish at the correct time so that the pearl fell appropriately into the box at the bottom of the sea. This exercise trained selective attention, visual perception, and visuomotor skills. All exercises were adapted to the patient’s performance by changing the difficulty and level at the beginning of each task. This was adapted by the therapist in charge. Patients received information about their performance during the training tasks. They could choose from a subset of exercises with the therapist, ensuring that all exercises were performed at a well-balanced frequency.

Figure 1. Computerized cognitive training exercises, representing the exercises of Freshminder; upper images: Doppelspiel (double play) and Perlenfische (pearlfish), bottom images: Schrittfolge (step sequence) and Aufgabenwechsel (tasks switching).

The VR group received treatment with an early version of Teora mind, conceptualized and developed by the medical device manufacturer Living Brain, a cognitive training in immersive VR. Teora mind is a certified medical device of risk class IIa according to Medical Device Regulation (EU) 2017/745. The certification procedure was executed by Notified Body TÜV Süd in 2021. The SG integrates ADL and cognitive training methods in a playful way to improve cognition and transfer into daily life and consists of realistic scenarios for training specific abilities.

The participants were trained in a virtual garden and kitchen scenario. Patients were asked to make coffee for multiple people in the kitchen. Therefore, they had to actively implement all the necessary steps for serving coffee, including filling the coffee machine with beans and water, changing the coffee filter,
choosing a mug, and considering whether the coffee had to be served with milk and sugar (Figure 2). Other exercises included sorting groceries according to their categories, for example, beverages, fruits and vegetables, and durable food for stock, or sorting out the refrigerator according to specifications, such as spoiled food. A time pressure component was added with increasing difficulty. All activities were implemented in real time and through patient interactions with the virtual environment.

**Figure 2.** Virtual reality training scenarios, upper images: virtual garden scenario, showing action planning board (left) and greenhouse with user harvesting cucumbers (right); bottom images: virtual kitchen scenario, showing user refilling coffee machine with coffee beans (left) and start screen when entering the kitchen (right).

In the garden scenario, patients were asked to plant seeds to grow strawberries, tomatoes, and cucumbers. In the growth phase, users had to ensure that there was always sufficient water in the bed and that popping weeds did not destroy the seedlings, which required watering the bed with a watering can. After successfully growing fruits and vegetables, users can produce juices from the harvested food. Participants had to plan their activity by first sorting picture cards describing each action step on a board in what they thought was the correct order. Next, they were required to confirm their orders. Finally, the system reported whether the order was correct or incorrect and indicated the steps that were arranged incorrectly. The exercise could begin only after the right order was achieved. The complexity varied depending on the difficulty level; at the highest level, planning was omitted and the activity had to be started immediately.

The main cognitive domains targeted by this VR training are EF, in particular action planning, problem-solving, selective attention, and working memory. The specific pieces of training themselves do not primarily focus on a single cognitive domain but rather on training the ADL, which typically comprises multiple cognitive domains simultaneously. Similar to the CCT intervention, the difficulty could be adapted to individual performance by changing to 20 different degrees of difficulty, which was adjusted by the therapist in charge. Patients received information on how they performed their training tasks in the VR application. In this trial, an early version, specifically designed for this clinical trial, was used. At that time, the medical device had not been introduced to the market. It was updated 2 times during the trial for minor bugs. VR devices of the type Oculus Quest (manufacturer: Meta) were used for the trial conduction. The device is a stand-alone HMD, and additional technical devices for experiencing immersive VR are not required. In addition to the VR device, the study nurses and therapists in charge received a tablet (Samsung Galaxy Tab A) for monitoring ongoing VR activities; all VR procedures were mirrored on the tablet in real time. This allowed patient-therapist interaction despite the high grade of immersion, particularly in the first training sessions of benefit. Interaction in immersive VR was enabled by the hand controller used by the patients. The blue hands visually represented them and were congruent with the actual movement of the upper extremities. The controllers included multiple buttons, of which only 2 were necessary to use the VR software. Moving the head makes it possible to gain a complete overview of the virtual world. Through a visible white beam, users can assess what they are currently pointing at in a virtual environment. This allowed for the selection of objects. To introduce the patients to the technology and virtual environment, they went through a dedicated tutorial at the beginning of each treatment. High resolution pictures of the training are available in Multimedia Appendices 1-4.
Procedure

In the first 2 sessions, which lasted approximately 60 minutes each, comprehensive testing (Alters-Konzentrations-Test [AKT], Wechsler Memory Scale—Revised [WMS-R], Trail Making Test [TMT] A and B, Tower of London—German version [TL-D], Short Form 36 [SF-36], visual analog scale [VAS, EQ-5D VAS], and Fragebogen zur Erfassung der Performance in VR [FEPVR]) was conducted (t₀). The testing consisted of psychological tests conducted by a psychologist and self-assessed questionnaires. After completing the testing, individuals were randomly assigned to 1 of the 2 training groups by drawing from an urn, which was performed by the study nurse. Only the responsible investigator had access to the list of individuals in each group. Because of the nature of the study, it was impossible to ensure the blindness of the investigator responsible for the intervention and participants. The process of participant randomization is described in the flowchart prepared according to the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Figure 3), and we used an analog randomization process by drawing notes from an urn. The whole CONSORT checklist is available in Multimedia Appendix 5. Simple randomization was performed by the study nurse. The treatment was scheduled for 18 to 25 sessions with a duration of 30 minutes for the individual, based on the date of admission and the participant’s immediate health and neurological condition. The sessions were conducted at least 3 to 5 days per week. Both groups received their allocated treatments from day 3 to day 20. A maximum of 25 treatment sessions were allowed.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

The study nurse instructed the intervention group on the device’s functions and use before the first VR training session. This included the general operation of the device from the outside with the attached buttons (on-off button and volume control), especially the handling of the hand controllers. After receiving the instructions, participants went through a VR tutorial, consisting of training in basic functionality, a memory game, and a recycling game, in which they were familiarized with general handling in the first training session. This tutorial demonstrated the interaction with the VR environment, virtual objects, and the handling of the entire experience. The tutorial was developed and tested with stroke patients, leading to high acceptance in a group of patients with stroke and healthy older adults, indicating that it is suitable for introducing this critical user group to immersive VR [16]. In addition, the familiarization phase helps reduce possible effects due to unawareness about the technology and negative attitudes toward it [29]. During all training sessions, the study nurse was able to track the process.
on a tablet and was thus able to provide accurate, customized assistance if needed.

The control group was also instructed on the operation of the used computer if needed, and an introduction to the program, explaining the different tasks, before the first training session was conducted by the study nurse in charge. During all training sessions, the study nurse was close to the participant and was thus able to provide accurate, customized assistance if needed.

The patients remained statically seated in the same place throughout all training sessions to minimize the potential risks of injury. Locomotion was performed using only a controller in a virtual environment. The study nurse documented the task results during the training sessions (ie, the time needed to finish a task and other observations), in a separate, pseudonymized document. The patients could choose from the exercises themselves, with the therapist ensuring that all exercises were performed at a well-balanced frequency. All participants could stop the treatment at any time, for example, if they felt unwell during training.

Immediately after the seventh training session, participants in both groups were asked to complete the Positive and Negative Affect Schedule (PANAS) questionnaire and rate their current subjective health status using the EQ-5D VAS. After the final treatment session, the test battery was repeated (t_1). Furthermore, a survey regarding the overall VR experience was conducted using open-ended questions, and the patient was asked for suggestions for improvement and satisfaction with the application.

**Dependent Measures**

**Overview**

Each participant was comprehensively tested at 2 points: before treatment (t_0) and immediately after the last session (t_1). The focus was on investigating the impact of the 2 cognitive training methods on diverse cognitive abilities. We tested the patients regarding working memory, selective attention, planning, attention, task organization, cognitive flexibility, and problem-solving. We also examined their quality of life, affect, and health status. In addition, a new measurement tool was tested to evaluate subjective performance in a situation similar to VR training to determine the extent to which the VR training was transferred to reality. The test battery included the following tests.

**AKT Measure**

The AKT [30] is a psychometric procedure that measures concentration ability and vigilance. The test was developed and standardized specifically for older people and is therefore adapted to their needs in terms of comprehensibility, task difficulty, and resilience requirements. Participants had to differentiate a figure at the top of the test sheet from a series of similar figures and cross it out. In this trial, we used the German version.

**TMT A and B Measure**

The TMT [31] is a cognitive test for measuring EF, particularly visuomotor abilities, which are essential for executing ADL properly [24]. In TMT A, patients had to connect numbers in ascending order; in TMT B, they had to connect numbers and letters in ascending and alternating order. The test has been standardized for patients with cognitive disorders undergoing neurorehabilitation and is often used for neurorehabilitation screening and diagnostics.

**TL-D Measure**

The TL-D is a transformation task that captures convergent problem-solving thinking and planning processes [32]. The objective of the task is to transfer 3 different-colored balls from an initial state to a specified target state in a minimum number of moves. The test consists of 20 tasks of varying severity and has been standardized in a sample of patients with neurological diseases; we used the German version of the test.

**WMS-R: Number Span Forward and Backward**

The WMS-R measures different memory functions [33]. In this study, a subset number span was used. First, the therapist reads a series of numbers of increasing length out loud. Depending on the task, the patient was then asked to recall the numbers in the correct order, either forward or backward. The WMS-R has been tested and standardized in patients with neurological and psychiatric disorders.

**SF-36 Measure**

The SF-36 questionnaire assesses patients’ health-related quality of life and consists of 36 items [34]. It measures 8 dimensions of subjective health: physical functioning, physical role functioning, physical pain, general health perception, vitality, social functioning, emotional role functioning, and psychological well-being. The SF-36 has been standardized in various patient populations; we used the German version of the instrument in this trial.

**EQ-5D VAS Measure**

Scientists use EQ-5D instruments developed by EuroQol for many different diseases. The EQ-5D VAS is a VAS in the standard layout of a vertical 20-cm scale with a range of values from 0 to 100, asking respondents to rate their current health status on the scale. The higher the rating, the better the indicated health status. The tool has been used in more than 117 countries and standardized for multiple patient groups.

**FEPVR Tool**

The VR performance instrument is a measurement tool to assess how well everyday activities can be performed and what effects training in VR has on daily performance. The test consisted of questionnaires, individual actions that were actively performed, and observation forms for therapists to rate patient activities. These activities included, among others, actively sorting laundry, planting strawberries, self-assessment of the ability to plan an action, and quantitative and qualitative feedback on the training experience. Our research group developed this tool specifically for this trial; however, it has yet to be validated in a large clinical sample. Therefore, the use was experimental. Book et al [35] and Graessel et al [36] developed and published a similar testing tool.
**PANAS Trial**

For this trial, we used the German version of the PANAS [37]. This questionnaire measured positive and negative affective states and traits. The patient rated the intensity of a sensation or feeling on a 5-point scale from “not at all” to “extremely.” Twenty items were presented.

**Power and Statistical Analysis**

Statistical analyses were performed using SPSS 29 (IBM Corp), with statistical significance set at $P<.05$. We performed repeated-measures ANOVA and post hoc tests to detect inner-subject and between-group changes between $t_0$ and $t_1$.

Comparing the VR and control groups in terms of cognitive performance and health-related measures involved group mean comparisons from 2 independent samples. Power calculations (power analysis tool GPower [Heinrich-Heine-Universität Düsseldorf, release 3.1.9.6 for Mac OS X]) revealed that our sample size of $n=21$ per group was sufficient to detect a standardized estimated population mean difference of $f=0.25$, which is considered a medium effect size, with a power of $1 − β = .88 − α = .05$. This shows that our small sample size did not constrain the detection of between-group effects at a conventional power goal of 0.8 with a medium effect size of $f=0.25$.

**Ethics Approval**

Approval for this pilot trial was obtained from the local Ethics Committee of the Regional Medical Association Hessen, Germany, under approval 2020-1768-fMP. The VR training Teora mind is a certified medical device (class IIa) in line with the European Medical Device Regulation and therefore obtained approval to conduct this trial from the state authority Bundesinstitut für Arzneimittel und Medizinprodukte under the file number 94208-5660-12543. All regulatory approvals were obtained before the recruitment of the first patient.

**Results**

**Baseline Characteristics**

The recruitment process started on November 1, 2020, and the last patient completed the treatment on March 31, 2022. We initially planned to include 52 patients but decided to close the study when we reached 42 patients finishing all treatment sessions because this sample size was large enough to calculate repeated-measures ANOVA. Further extension of the trial was not possible due to administrative reasons. A total of 21 patients received the novel VR treatment, whereas the other 21 received control CCT. Furthermore, 33% (14/42) of the patients were female. Patients in the control group were, on average, aged 67.3 (SD 4.6) years at trial participation, while the VR group had a mean age of 68.3 (SD 14.5) years at trial participation; no statistical difference between both groups regarding age was found ($P=.69$). Detailed information on the brain damage of the included participants is found in Table 1. For more detailed information on the demographic characteristics, mean performance, and SD for all sample measures in this clinical trial, please refer to Tables 2 and 3.

Of the 642 patients receiving treatment at the center at the time of trial conduction, 73 patients were primarily included for participation, of which 42 (14 female patients) completed all treatment sessions. However, 31 patients were not able to complete the minimum treatment number for the following reasons: COVID-19 ($n=4$), early dismissal ($n=12$), psychiatric comorbidities not detected earlier ($n=3$), and treatment discontinuation due to a high incidence of COVID-19 cases at the trial center ($n=12$).

Of the 42 included patients, 17 (28%) had a right-sided pathology, 22 (52%) had left-sided pathology, and 3 (7%) had damage occurring in both brain hemispheres.

Table 1. Types of brain damage.

<table>
<thead>
<tr>
<th>Brain region</th>
<th>Patients, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebral infarction</td>
<td>33</td>
</tr>
<tr>
<td>Brain hemorrhage</td>
<td>3</td>
</tr>
<tr>
<td>Subarachnoid hemorrhage</td>
<td>2</td>
</tr>
<tr>
<td>Basal ganglia hemorrhage</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2. Demographic characteristics of control and VR group.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control group (n=21)</th>
<th>VR patients (n=21)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>67.3 (4.6; 59-74)</td>
<td>68.3 (14.5; 36-93)</td>
<td>.69</td>
</tr>
<tr>
<td>MMSTb, mean (SD; range)</td>
<td>26.95 (2.04; 24-30)</td>
<td>27.05 (2.11; 24-30)</td>
<td>.97</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>.67</td>
</tr>
<tr>
<td>Female</td>
<td>8 (38)</td>
<td>6 (29)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13 (62)</td>
<td>15 (71)</td>
<td></td>
</tr>
</tbody>
</table>

aVR: virtual reality.
bMMST: Mini Mental State Test.
Table 3. Mean performance and SD in the control and VRa group for all measures used in the trial at baseline.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control group, mean (SD; range)</th>
<th>VR patients, mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT(^a) A</td>
<td>93.7 (47.83; 36-205)</td>
<td>80.95 (47.84; 39-200)</td>
</tr>
<tr>
<td>TMT B</td>
<td>326.9 (216.7; 91-853)</td>
<td>29.2 (244.3; 96-1056)</td>
</tr>
<tr>
<td>TL-D(^b)</td>
<td>54.9 (31; 3-99)</td>
<td>42.62 (32.1; 1-99)</td>
</tr>
<tr>
<td>EQ-5D VAS(^c)</td>
<td>45.5 (21.02; 0-80)</td>
<td>49.05 (17.22; 0-75)</td>
</tr>
<tr>
<td>AKT_T_PR(^d)</td>
<td>43.04 (21.4; 0-74.4)</td>
<td>41.07 (21.8; 0-97)</td>
</tr>
<tr>
<td>WMS-R(^e) forward</td>
<td>39.6 (29.23; 0-98)</td>
<td>41.9 (34.99; 0-98)</td>
</tr>
<tr>
<td>WMS-R backward</td>
<td>28.2 (29.82; 0-98)</td>
<td>36.7 (32.96; 2-97)</td>
</tr>
<tr>
<td>SF-36(^f) energy</td>
<td>42.83 (20.47; 0-80)</td>
<td>42.38 (20.77; 5-80)</td>
</tr>
<tr>
<td>SF-36 emotional well-being</td>
<td>55.15 (22.65; 0-88)</td>
<td>64.24 (23.94; 16-100)</td>
</tr>
<tr>
<td>SF-36 social functioning</td>
<td>36.87 (26.74; 0-100)</td>
<td>50.6 (31.99; 0-100)</td>
</tr>
<tr>
<td>SF-36 pain</td>
<td>51.75 (34.57; 0-100)</td>
<td>63.88 (34.6; 0-100)</td>
</tr>
<tr>
<td>SF-36 general health</td>
<td>47.5 (14.18; 25-75)</td>
<td>46.85 (13.82; 10-70)</td>
</tr>
<tr>
<td>SF-36 physical function</td>
<td>29.75 (30.45; 0-100)</td>
<td>20.77 (20.08; 0-66.6)</td>
</tr>
<tr>
<td>SF-36 role limitation in physical health</td>
<td>22.37 (36.22; 0-100)</td>
<td>13.49 (30.44; 0-100)</td>
</tr>
<tr>
<td>SF-36 role limitation emotional problems</td>
<td>35.09 (45.1; 0-100)</td>
<td>31.66 (43.9; 0-100)</td>
</tr>
</tbody>
</table>

\(^{a}\)TMT: Trail Making Test.  
\(^{b}\)TL-D: Tower of London—German version.  
\(^{c}\)VAS: visual analog scale.  
\(^{d}\)AKT_T_PR: Alters-Konzentrations-Test time percentile rank.  
\(^{e}\)WMS-R: Wechsler Memory Scale—Revised.  
\(^{f}\)SF-36: Short Form 36.

**Nausea and Motion Sickness**

Of the 21 patients in the VR group, none reported problems with motion sickness or associated nausea, and no treatment had to be stopped due to motion sickness or nausea. This is congruent with the results of our feasibility trial with patients with stroke and healthy older adults, where we demonstrated that applying specific principles in VR development, such as having complete control of all motions in VR, reduces the risk of experiencing motion sickness and is feasible for use in special needs groups such as patients with stroke [16]. Patients in the control group using CCT reported no negative side effects. The patients conducted, on average, 18.7 (SD 0.99) successful treatment sessions.

**Treatment Satisfaction**

Patients were asked as part of the FEPVR tool to rate the treatment satisfaction on a Likert scale ranging from 1 (very dissatisfied) to 7 (very satisfied). Both groups reported a tendency toward high acceptance and satisfaction with both treatments (M=5.7). The treatment also created the impression that it helped against the present condition in both groups (VR: M=5.5, CCT: M=5.7), rated from 1 (I completely disagree) to 7 (I fully agree).

**Adverse Events**

While the trial was conducted, no severe adverse event or adverse event was linked to the intervention in either group; nevertheless, we had to report adverse events as the clinical trial was conducted amid the COVID-19 pandemic in 2020 and 2021. Owing to the general incidence of infections in Germany, 4 patients included in the trial were infected with SARS-CoV-2 and were isolated and excluded from the study.

**Survey Completion**

Of the patients who completed the study, 97% (41/42) of the participants completed all the questionnaires and tests, with only occasional missing items or tests. This might be explained by the fact that all surveys were conducted in the presence of a therapist or neuropsychologist working and trained in the clinic.

**Cognitive Tests**

To evaluate the impact of both treatments on cognitive status, subjective health, and quality of life, we used SPSS Statistics (version 29, IBM) and calculated repeated-measures ANOVA. For the comparisons between the 2 groups between the pre- and posttest, repeated-measures ANOVA and Bonferroni post hoc test indicated that TL-D scores significantly differed in the VR group after the intervention (P=.046). Table 4 presents the results of all other conducted tests. As a nonsignificant trend emerged in the TMT A (main effect: P=.01; time by group interaction: P=.09) and WMS-R (main effect: P=.006; time by group interaction: P=.18) results, a post hoc test was also performed. No further post hoc tests were performed for the other variables.
Table 4. Clinical parameters of the virtual reality (VR) and computerized cognitive training groups before and after the intervention.

<table>
<thead>
<tr>
<th>Measure and time point</th>
<th>Control group(^a) (n=21)</th>
<th>VR group(^b) (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TMT</strong> A scores, mean (SD; range)(^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>93.7 (47.83; 36-205)</td>
<td>80.95 (47.84; 39-200)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>89.95 (55.2; 26-257)</td>
<td>61 (22.87; 32-135)</td>
</tr>
<tr>
<td><strong>TMT B scores, mean (SD; range)(^e)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>326.9 (216.7; 91-853)</td>
<td>292.24 (244.26; 96-1056)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>283.75 (198.86; 64-877)</td>
<td>268.29 (217.6; 99-855)</td>
</tr>
<tr>
<td><strong>WMS-R(^f), mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>39.6 (29.23; 0-98)</td>
<td>41.9 (34.99; 0-98)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>44.75 (31.24; 0-88)</td>
<td>55.95 (30.29; 0-98)</td>
</tr>
<tr>
<td><strong>TL-D(^h) scores, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>54.9 (31; 3-99)</td>
<td>42.62 (32.1; 1-99)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>52.2 (33.7; 2-99)</td>
<td>73.1 (32.5; 7-99)</td>
</tr>
<tr>
<td><strong>EQ-5D VAS(^i) scores, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>45.5 (21.55; 0-80)</td>
<td>49.05 (17.22; 0-75)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>46.84 (25.01; 0-80)</td>
<td>62.86 (21.3; 10-95)</td>
</tr>
<tr>
<td><strong>AKT(^j) scores, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>87.47 (62.98; 32-281)</td>
<td>89.23 (69.57; 28-357)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>93.15 (67.83; 33-351)</td>
<td>87.86 (69.56; 35-350)</td>
</tr>
<tr>
<td><strong>SF-36</strong> energy domain, mean (SD; range)(^o)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>42.83 (20.47; 0-80)</td>
<td>42.38 (20.77; 5-80)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>51.32 (20.27; 15-90)</td>
<td>47.81 (21.3; 10-85)</td>
</tr>
<tr>
<td><strong>SF-36 emotional well-being domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>55.15 (22.65; 0-88)</td>
<td>64.24 (23.94; 16-100)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>64.63 (17.55; 32-100)</td>
<td>66.1 (22.36; 24-100)</td>
</tr>
<tr>
<td><strong>SF-36 social functioning domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>36.87 (26.74; 0-100)</td>
<td>50.6 (31.99; 0-100)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>48.13 (28.76; 0-100)</td>
<td>54.17 (34.08; 0-100)</td>
</tr>
<tr>
<td><strong>SF-36 pain domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>51.75 (34.57; 0-100)</td>
<td>63.88 (34.6; 0-100)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>59.38 (30.27; 20-100)</td>
<td>60.48 (33.3; 0-100)</td>
</tr>
<tr>
<td><strong>SF-36 general health domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>47.5 (14.2; 25-75)</td>
<td>46.85 (13.82; 10-70)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>48.25 (16.8; 10-90)</td>
<td>48.57 (16.52; 20-90)</td>
</tr>
<tr>
<td><strong>SF-36 physical functioning domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>29.75 (30.46; 0-100)</td>
<td>20.77 (20.07; 0-66.7)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>27.89 (21.99; 0-65)</td>
<td>36.43 (29.33; 0-95)</td>
</tr>
<tr>
<td><strong>SF-36 role limitation physical health domain, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>22.37 (36.22; 0-100)</td>
<td>13.49 (30.44; 0-100)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>21.25 (35.61; 0-100)</td>
<td>15 (31.83; 0-100)</td>
</tr>
<tr>
<td><strong>SF-36 role limitation emotional problems, mean (SD; range)</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Measure and time point

<table>
<thead>
<tr>
<th>Measure and time point</th>
<th>Control group(^a) (n=21)</th>
<th>VR group(^b) (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretreatment</td>
<td>35.09 (45.1; 0-100)</td>
<td>31.66 (43.9; 0-100)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>45.33 (49.68; 0-100)</td>
<td>42.86 (48.47; 0-100)</td>
</tr>
</tbody>
</table>

\(^a\)Control group using computerized cognitive training.
\(^b\)VR group using immersive VR cognitive training.
\(^c\)TMT: Trail Making Test.
\(^d\)Repeated-measures ANOVA: no significant interaction detected (P=.09).
\(^e\)Repeated-measures ANOVA: no significant interaction detected (P=.68).
\(^f\)WMS-R: Wechsler Memory Scale—Revised.
\(^g\)Repeats measures analysis of variance: no significant interaction detected (P=.18).
\(^h\)TL-D: Tower of London.
\(^i\)Repeated-measures ANOVA: time, \(F_{1,39}=4.255; P=.046\) (significant); time \& group, \(F_{1,39}=6.07; P=.02\) (significant). A Bonferroni post hoc test was significant between the VR group at pretest and the VR group at posttest (\(P_{\text{bonf}}=.02\); \(\eta^2=0.135\)).
\(^j\)VAS: visual analog scale.
\(^k\)Repeated-measures ANOVA: no significant interaction detected (P=.16).
\(^l\)AKT: Alters-Konzentrations-Test.
\(^m\)Repeated-measures ANOVA: no significant interaction detected (P=.52).
\(^n\)SF-36: Short Form 36.
\(^o\)Repeated-measures ANOVA: no significant interaction detected (\(P_{\text{pain}}=.28\); \(P_{\text{energy}}=.69\); \(P_{\text{emo_wb}}=.31\); \(P_{\text{role_lim_phys}}=.59\); \(P_{\text{role_lim_emo}}=.87\); \(P_{\text{physfunc}}=.06\); \(P_{\text{gen_health}}=.18\); \(P_{\text{emofunc}}=.18\); \(P_{\text{physfunc}}=.06\); \(P_{\text{role_lim_phys}}=.88\); \(P_{\text{role_lim_emo}}=.69\)).

### PANAS Trial

Patients were asked to describe their affects on the allocated treatment after 7 sessions of trial participation. Affect was measured with the German version of PANAS, and patients were asked to rate their affect immediately after treatment. An ANOVA was performed, resulting in \(F_{1,36}=0.879\) (\(P=.63\)) for the positive affect scale and \(F_{1,36}=0.954\) (\(P=.57\)) for the negative affect scale. The test results showed no difference between the intervention and control groups regarding positive and negative affect toward the kind of treatment received by both groups during the trial. Both groups reported high positive affect, with the VR group at mean 29.809 (SD 7.98) and the control group at mean 25.6 (SD 8.23), showing a slightly higher positive affect in favor of the VR group. Negative affect results were rated mean 15.95 (SD 8.59) in the VR group and mean 15.66 (SD 6.15) in the control group, showing similar negative affect ratings in both groups.

### Summary

The critical effects for both groups are presented in Table 4, showing mean changes between \(t_0\) and \(t_1\) and the results of repeated-measures ANOVA for group effects. Figure 4 shows the mean changes for TL-D results in both groups.

**Figure 4.** Results of repeated measures analysis of variance Tower of London—German version (TL-D). CCT: computerized cognitive training; VR: virtual reality.
Discussion

Overview

To our knowledge, this is one of the few trials comparing immersive VR cognitive training, incorporating ADL and SG and focusing on the ability to perform ADL, with a multitude of cognitive domains and not single domains only, with CCT. In this randomized controlled trial, we aimed to determine the impact of both cognitive training in immersive VR and CCT on cognition and the ability to perform ADL in patients in inpatient neurological rehabilitation with an MMST score >20 in rehabilitation phases C and D, combining immersive VR with paradigms of neuropsychology and occupational therapy. In the patient group receiving VR therapy, we found a statistically significant improvement in executive functioning within the 4-week treatment interval. Some nonsignificant trends were detected for attention and processing speed, working memory, and subjective health. No significant changes were observed in the control group.

In recent years, VR systems found their way to neurological rehabilitation. However, being used most often for motor rehabilitation, only a few systems have been developed and evaluated concerning their impact on cognition in patients with acute brain affection. Furthermore, some developed trainings have only been used for research purposes but have yet to undergo more extensive testing or application in clinical practice. In addition, several cognitive trainings in VR were developed to target a specific cognitive domain, for example, memory [38] and attention [39].

Principal Findings

We compared the VR group with a control group using CCT, which represents care as usual in German rehabilitation facilities. Both groups were able to engage in their assigned interventions successfully. Despite our initial concerns regarding the number of sessions to be conducted in VR, no significant side effects were reported. All patients participated in all sessions required to complete the trial. Some initial hurdles had to be overcome, for example, mainly introducing older adult patients to an entirely new technology or ensuring continuous disinfection of all equipment, as the trial took place at a time when the COVID-19 pandemic situation in Germany was very tense. However, after the first treatment session, those problems were solved without further external intervention. Overall, the acceptance and usability of both technologies were observably high, which, in terms of the VR experience, is congruent with the results of our feasibility study using an earlier version of the software used in this trial [16].

The demographics between the 2 groups were similar: not significantly different in age and pathology but with a larger proportion of male participants in the trial. The trial’s sample size was too small to make far-reaching statements, but clear tendencies provided the first indication of training effects that opened up room for further research in a larger trial.

We detected the effects of immersive VR on different cognitive functions in patients with acute brain affection receiving treatment during inpatient neurological rehabilitation. Comparing both groups by calculating repeated-measures ANOVA, we detected improvements in some cognitive domains in the VR group but no significant improvement in any cognitive function in the control group. By conducting the TL-D, specifically assessing problem-solving and planning, we found significant improvements from pre- to postintervention phase in the VR group, and the Bonferroni post hoc test also verified these significant improvements with a medium-sized effect of $\eta^2=0.135$. Oliveira et al [40] conducted a trial to investigate the impact of VR on ADL in stroke rehabilitation. They found that VR training resulted in better tests measuring EF, global cognition, attention, and memory training. This partly overlaps with the results of this study. We also observed improvements in attention, processing speed, and subjective health status in the VR group. However, these did not become significant in the repeated-measures ANOVA, while the scores for the named variables persisted at the $t_p$ level after the treatment in the control group. No significant improvements and differences between groups were detected in concentration and memory in both groups.

The improvements in subjective health state are of interest for further research, as Domínguez-Tellez et al [41] ascertained in their systematic review that VR interventions for upper limb rehabilitation were a part of improving quality of life. They hypothesized that an improvement in the actual pathology leads to better abilities for carrying out ADL, and an improvement in quality of life goes hand in hand with a higher grade of participation. The higher levels of subjective health state in the VR group (mean 62.86, SD 21.03) compared with the control group (mean 46.84, SD 25.01) after the intervention, with an improvement on the EQ-5D VAS-scale of 13.81 points in the VR group and 1.34 points in the control group between $t_0$ and $t_1$, although not significantly different, showed a tendency toward the impact of VR treatment when considering the significant improvement in TL-D scores in the VR group and the resulting implications for participation in daily life. More thorough research is necessary to better understand the interrelationship between these constructs and the results of the SF-36 subscales.

Looking at the results the VR group achieved in planning and problem-solving (TL-D), we would have expected to notice a significant difference in TMT results as well, as both tests are positively correlated [42]. TMT specifically measures cognitive flexibility, alternating attention, sequencing, visual search, and motor speed, which overlap with the measured domains of TL-D, predominantly measuring planning and problem-solving. The missing significant improvement in the abilities measured by TMT in this trial raises the question of how specifically the VR training stimulates particular cognitive skills. We still see a tendency for improvement in these cognitive abilities, as the VR group achieved more improvement between $t_0$ and $t_1$ than the control group, as measured by TMT. A. We found a nonsignificant trend in the repeated-measures ANOVA ($P=0.9$).

As we expected a correlation between TL-D and TMT results, we calculated a Bonferroni post hoc test, which resulted in a significant improvement in the VR group (Bonferroni $P=0.02$). This is only a trend, which cannot be used to draw far-reaching
conclusions. Nevertheless, it is interesting, as this might correlate with the significant improvement in TL-D measures in the same group. Nevertheless, it would be interesting to investigate in a larger trial if an optimized version of the training software leads to more diversified and measurable effects.

In addition, other studies cited in this paper have reported further improvements in multiple cognitive subdomains. In our study, significant improvements were found only in the EF. One explanation for the difference in results is that both groups received only 1 therapy to treat cognitive dysfunction rather than simply augmenting the existing treatment. A 2017 Cochrane Review found that in many studies, VR therapy was added to traditional treatment, resulting in a higher total amount of therapy [43]. Therefore, it cannot be said with certainty whether VR therapy alone is causal for the improvement in these studies or whether the overall increased measure of treatment compared with the control group is the determining factor. The fact that each group in our study received only 1 cognitive rehabilitation treatment means that the effects can be more precisely attributed to a single treatment method.

When developing the software used in collaboration with occupational therapists, neuropsychologists, neurologists, and patients, we were particularly faced with the lack of proximity to daily life of current CCT, which led us to the question of how we could depict everyday life and its challenges in a way that is easy to understand, therapeutically beneficial, and motivating, so that patients could adhere to training over a more extended period. The results showed the VR training used in this trial. The following components have played a significant role in its development and could provide a directional indicator for further development of new training methods for cognitive rehabilitation.

The results of this trial helped us formulate hypotheses on why VR-based training might lead to better results in cognitive rehabilitation than PC-based interventions:

**Ecological Validity**

In rehabilitation, particularly when focusing on returning to daily life, it is vital to have ecologically valid training simulations, for which VR is well suited [40,44]. In our application, we presented a garden and kitchen scenario to simulate activities and situations in which subjects found themselves after rehabilitation. In reality, as it is either dangerous or unavailable, it is often not viable to let patients train in real-life settings. VR enables stimulation with settings as close as possible to reality, potentially facilitating a more effortless transfer from training to reality.

The results of this trial indicate that high ecological validity is beneficial for training cognition. Patients in the VR group achieved significantly higher scores in planning and problem-solving than their counterparts in the control group. In addition, they reported better scores in terms of speed and health status than the conventionally treated patients. This is congruent with the results of Faria et al [21], who examined the benefits of immersive ADL simulation in neurological rehabilitation [21].

**Gamification and Immersion**

Gamification adds games or game-like elements to encourage participation or engagement. Engagement is a crucial component of rehabilitation; in patients with stroke, it increases the rate of rehabilitation and rehabilitation outcomes of survivors of stroke [45]. The VR experience relies to a great extent on the grade of immersion. Immersion describes the effect of a virtual environment that causes the user’s awareness of being exposed to illusory stimuli to fade into the background to such an extent that the virtual world is perceived as real. High-grade immersion can be created using HMD-VR. The user wears device-like goggles on the head over the eyes, effectively masking the actual environment.

Combining these 2 aspects and adding up the feature of receiving immediate feedback from the virtual environment on performance leads to increased motivation and, subsequently, patients stick with the treatment longer [43]. Yoshida et al [46] discussed motivation in patients with subacute stroke and discovered that it is mainly influenced by extrinsic reward factors (eg, positive feedback provided by therapists). However, if basic psychological needs are met, for example, autonomy or competence, which play a critical role in stroke rehabilitation, the motivation pattern switches from extrinsic to intrinsic [46]. Long after an acute event, improvements in rehabilitation can still be achieved, and patients’ self-perception regarding the improvements is paramount.

**On the Right Dosage**

Participants in this trial received a treatment dosage of 18.7 sessions over 4 to 6 weeks, each lasting 30 minutes (9.35 hours per participant). Other trials reported an average intervention duration of 13.3 hours [39,40,47-50], ranging from 4 to 24 hours, with a session duration of 30 to 60 minutes, 3 to 5 times a week. Our dosing was sufficient to achieve a measurable effect in some domains and was also oriented toward the care reality in German rehabilitation hospitals. To attain more significant outcomes in multiple domains, it would be interesting to see if other training tasks in immersive VR create an effect on cognition and if a varying number of treatment sessions beyond inpatient care leads to different results than those reported here. On the basis of the results of this trial, we conclude that 18 to 19 sessions over 30 minutes are sufficient to achieve a measurable treatment effect. Still, we cannot determine whether fewer sessions would create similar effects if more treatment automatically leads to more significant outcomes and if those effects vary between patient groups. More extensive trials with blinding, subgroups, and longer treatment duration are necessary to answer this question.

**Transfer to Daily Life and How to Measure it**

One end point of this trial, which we were not able to meet, was to identify whether the treatments in both groups induced transfer effects on daily life by examining if patients in either group were in a better state to perform instrumental ADL, such as preparing food or handling the washing machine, after the intervention. While designing this trial, we noted that standardized neuropsychological testing batteries are well suited to test particular cognitive domains and performance therein,
but that daily life is more complex—it can not be reduced to 1 or 2 cognitive processes at the same time—and that functional capabilities of managing everyday life as the end goal of rehabilitation are not sufficiently represented by those testing batteries when focusing on cognitive rehabilitation [40]. Therefore, we developed a testing tool, FEPVR. It was used for the first time in this trial, and we faced some issues with properly capturing the ADL. We plan to publish the battery and its manual in a revised version, including experiences from use in this trial so that further development and testing by other research groups is possible.

**Limitations**

Although this study provides valuable insights into the impact of immersive VR on cognitive rehabilitation, several questions remain unanswered. Specifically, we identified the following limitations that should be considered: future research in this field should be conducted with larger sample groups. This randomized controlled pilot study builds a profound understanding of the potential immersive VR can have in neurological rehabilitation for cognitive impairment in patients with MMST scores >20, affected by stroke, to the largest extent. Conclusions cannot be drawn for patients below this cut-off score. We also need more information on the manifestation of training effects, for example, if the measured results are still the same 6 to 12 months after the intervention and if an actual transfer effect to ADL occurs. Therefore, long-term studies with follow-up over multiple years are necessary.

The applied VR environment was limited to only 2 training scenarios inspired by ADL. More training scenarios replicating other ADL should be developed and tested in future studies.

Even though there are findings about the application of VR in rehabilitation of other patient groups, for example, multiple sclerosis, psychiatric disorders, and pain management [51], the application used here has chiefly been tested in the rehabilitation of patients with acute brain affection. Therefore, the evaluation needs to consider patients from the abovementioned groups to draw further conclusions in other patient groups.

**Conclusions and Outlook**

Our group of psychologists, software developers, 3D artists, neuropsychologists, and occupational therapists has developed a new form of immersive VR-based cognitive rehabilitation training. After proving the feasibility and acceptability in stroke patients with stroke, we demonstrated the initial therapeutic effects of a novel immersive VR training on EF in patients undergoing inpatient neurological rehabilitation. Data were collected over 18 months, including multiple recruitment pauses due to the COVID-19 pandemic between 2020 and 2022. The supervision of patients was necessary for conducting treatment, but it was administered by trained therapy assistants, freeing time for neuropsychologists and occupational therapists. None of the patients in the VR group had previously been exposed to VR, but that did not diminish their excitement about receiving the experimental treatment and getting the most out of it.

It was reported earlier that combining cognitive training with physical activity leads to even better results; for example, a group of older adults with mild cognitive impairment showed more significant improvements in multiple cognitive domains when receiving combined therapy than physical or cognitive training alone [52]. Investigating whether the treatment effects of VR-based cognitive training can be increased by additional full-body physical training would be an interesting research topic for future trials. We also intend to depict everyday life in VR even more with new training scenarios, a higher level of gamification, and using techniques to personalize the treatment itself to a greater extent.

Our goal is to ensure that treatments in neurorehabilitation are improved and that all parties involved profit from new developments. The excellent acceptance, low side effects, and initial therapeutic effects of immersive VR treatment presented here offer new therapeutic options, for example, applying SG in routine therapy [53]. The results of this pilot trial show that a short but intense period of treatment leads to measurable improvement in EF, a manifold construct consisting of behavioral and higher-order cognitive skills necessary for independent functioning and carrying out everyday life, such as shopping, managing money, and preparing meals [3,8]. It would be interesting to determine if these results can be extended if patients continue treatment at home and increase the overall number of treatment sessions.

Temehy et al [4] explored the needs of patients with stroke after discharge from rehabilitation; they discovered that patients often not only need physical rehabilitation but also psychological services, comprising cognitive rehabilitation. Shipley et al [54] cite a patient who says, “it should have been a formal process to gain access to a psychologist, and it would have been beneficial from an earlier time point, such as from the acute hospital.” Access to qualified and trained neuropsychologists in Germany is difficult, as the Gemeinsamer Bundesausschuss (Federal Joint Committee) states that only 1100 neuropsychologists in outpatient care are available. The 2019 German Society for Neuropsychology list contains even fewer specialists with only 350 practitioners in outpatient settings [55]. This shortcoming highlights the need for complementary treatment approaches. This is in line with what the COVID-19 pandemic has uncovered: the need for digitized rehabilitation tools, bringing the general problem of accessibility of rehabilitation services to center, for example, for people living in rural areas or simply an insufficient number of neuropsychologists or occupational therapists in Germany, which prevents the actual implementation of necessary aftercare in neurological rehabilitation with a regular high number of treatment sessions for manifesting rehabilitation progress. Further development of video game–based rehabilitation approaches with digitized platforms to bring care home should be the focus of other groups active in this field, which is also emphasized by Ong et al [56] and Salisbury et al [57], who stated that VR is a promising technology to foster home care and shorten hospitalization. As prices for HMD-VR devices continuously decrease and the acceptance of digitized treatments in Germany increases, it augments the chance that access to VR-assisted rehabilitation will become more accessible and a standard in a wide range of treatment settings.
Acknowledgments
The Federal State of Baden-Württemberg supported this work through the Innovationsgutschein BW High-Tech Startup funding program.

Data Availability
The data sets generated during this study are available from the corresponding author only upon reasonable request.

Conflicts of Interest
This work was partially funded by the medical device manufacturer Living Brain GmbH, Heidelberg, Germany, who developed the used virtual reality software. The coauthors of this publication, BS and JS, are cofounders of Living Brain. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Greenhouse screenshot 1.
[ PNG File, 691 KB - games_v11i1e45816_app1.png ]

Multimedia Appendix 2
Greenhouse screenshot 2.
[ PNG File, 3493 KB - games_v11i1e45816_app2.png ]

Multimedia Appendix 3
Kitchen 1.
[ PNG File, 489 KB - games_v11i1e45816_app3.png ]

Multimedia Appendix 4
Small kitchen lobby.
[ PNG File, 948 KB - games_v11i1e45816_app4.png ]

Multimedia Appendix 5
CONSORT-EHEALTH checklist (V. 1.6.1).
[ PDF File (Adobe PDF File), 1375 KB - games_v11i1e45816_app5.pdf ]

References


Abbreviations

ADL: activities of daily living  
AKT: Alters-Konzentrations-Test  
CCT: computerized cognitive training  
CONSORT: Consolidated Standards of Reporting Trials  
EF: executive functions  
FEPVR: Fragebogen zur Erfassung der Performance in VR  
HMD: head-mounted display  
MMST: Mini Mental Status Test  
PANAS: Positive and Negative Affect Schedule  
SF-36: Short Form 36  
SG: serious game  
TL-D: Tower of London—German version  
TMT: Trail Making Test  
VAS: visual analog scale  
VR: virtual reality  
WMS-R: Wechsler Memory Scale—Revised

Please cite as:  
Specht J, Stegmann B, Gross H, Krakow K  
Cognitive Training With Head-Mounted Display Virtual Reality in Neurorehabilitation: Pilot Randomized Controlled Trial  
JMIR Serious Games 2023;11:e45816  
URL: https://games.jmir.org/2023/1/e45816  
doi:10.2196/45816  
PMID:37477957

©Julian Specht, Barbara Stegmann, Hanna Gross, Karsten Krakow. Originally published in JMIR Serious Games (https://games.jmir.org), 21.07.2023. 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.
Original Paper

Design Preferences for a Serious Game–Based Cognitive Assessment of Older Adults in Prison: Thematic Analysis

Rhys Mantell1, MPH; Adrienne Withall1,2,3, PhD; Kylie Radford2,3,4, PhD; Michael Kasumovic5,6, PhD; Lauren Monds7, PhD; Ye In Jane Hwang1,2, PhD

1School of Population Health, University of New South Wales, Sydney, Australia
2Ageing Futures Institute, University of New South Wales, Sydney, Australia
3School of Psychology, University of New South Wales, Sydney, Australia
4Neuroscience Research Australia, Sydney, Australia
5School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, Australia
6Arludo, Sydney, Australia
7Addiction Medicine, Central Clinical School, The University of Sydney, Sydney, Australia

Corresponding Author:
Rhys Mantell, MPH
School of Population Health
University of New South Wales
Samuels Building, F25, Samuel Terry Avenue, Kensington
Sydney, 2052
Australia
Phone: 61 0427161651
Email: r.mantell@unsw.edu.au

Abstract

Background: Serious games have the potential to transform the field of cognitive assessment. The use of serious game–based cognitive assessments in prison environments is particularly exciting. This is because interventions are urgently needed to address the rapid increase in the number of currently incarcerated older adults globally and because of the heightened risks of dementia and cognitive decline present in this population. Game-based assessments are assumed to be fun, engaging, and suitable alternatives to traditional cognitive testing, but these assumptions remain mostly untested in older adults. This is especially true for older adults in prison, whose preferences and needs are seldom heard and may deviate from those previously captured in studies on cognition and serious games.

Objective: This study aimed to understand the design preferences of older adults in prison for a game-based cognitive assessment.

Methods: This study used reflexive thematic analysis, underpinned by critical realism, and applied the technique of abduction. Overall, 4 focus groups with a total of 20 participants were conducted with older adults (aged ≥50 years; aged ≥45 years for Aboriginal and Torres Strait Islander people) across 3 distinct prison environments in Australia.

Results: Self-determination theory was used as a theoretical foundation to interpret the results. Overall, 3 themes were generated: Goldilocks—getting gameplay difficulty just right through optimal challenge (the first theme emphasizes the participants’ collective desire for an individualized optimal level of difficulty in serious gameplay), Avoiding Childish Graphics—gimmicky gameplay can be condescending (the second theme raises the importance of avoiding immature and childlike gameplay features, as some older end users in prison felt that these can be condescending), and A Balanced Diet—meaningful choice and variety keeps game-based assessments fun (the third theme highlights the strong user preference for meaningful choice and variety in any serious game–based cognitive assessment to maximize in-game autonomy).

Conclusions: The collection of these themes provides novel insights into key game design preferences of marginalized older adults.

(JMIR Serious Games 2023;11:e45467) doi:10.2196/45467

KEYWORDS

serious game; gamification; cognitive assessment; prison; older adults; older prisoners; game design; self-determination theory
Introduction

Assessing Cognition Through Serious Games

Gamification stands as one of the most promising solutions to enhance the user experience of traditionally monotonous tasks [1]. The process of gamification includes applying game design elements such as challenges, scoring, graphics, and narratives to nongame environments (e.g., psychometric testing) to increase engagement and motivation [2]. When gamified tasks are embedded into an immersive user experience, they are typically referred to as serious games. Serious games present users with unique and exciting environments to complete objectives that go beyond entertainment [3]. An emerging area of research is the study of serious games for cognitive assessment. Although there are well-evidenced benefits of using traditional cognitive assessment tools (proven psychometric validity, high sensitivity to neurological disease, etc) [4-6], they are often seen as boring and repetitive by users [7]. This can reduce task engagement [2] and affect the reliability of the performance data collected [1,8]. People also report test anxiety and self-stigma regarding low literacy or education levels when performing traditional cognitive tests [9-12]. In addition, there is evidence that some widely used neuropsychological tests are not culturally safe [13-15].

Serious games have the potential to transform the field of cognitive assessment. Game-based assessments appear to be suitable alternatives to traditional tasks given their potential to improve motivation and enjoyment and support individuals to produce their best effort, without making them feel overly bored, anxious, or distracted [1,16]. Recent studies have indicated that the experience of playing digital cognitive games was perceived as less stressful, more interesting, and more enjoyable when compared with standard cognitive tasks [2,3]. Serious games can also offer a brief, cost-effective, and scalable substitute for more traditional assessment methods [8,17]. In addition, serious games might support more ecologically valid assessment through realistic context and gameplay by engaging cognitive processes in a way that is more similar to a real-life situation [18].

Although the user benefits of enhanced motivation, engagement, and enjoyment are promising, they are not automatically derived through gamification. In recent years, some gamified cognitive assessments have assumed that, because something is presented as a game, it is fun. In a systematic review (conducted in 2021) of game-based cognitive assessments [1], less than half of the games included in the study evaluated any aspect of gameplay or user preference. Furthermore, although many studies have reported the positive effects of serious games on enjoyment [8], additional findings suggest that some games do not increase engagement, with Birk et al [19] concluding that their gamification actually reduced task engagement when compared with a traditional task. These findings suggest that serious games that assume user enjoyment and engagement overlook the complexity of end-user preferences. Poorly designed and executed serious games can be expensive and ineffective and are unlikely to result in a more enjoyable and motivating experience than traditional tasks [1,16,19,20]. If games are to be used as cognitive assessment tools, they must first be designed with the end user’s enjoyment and motivation in mind. Furthermore, ensuring that a game-based assessment is culturally safe to play is critical [21]. This means acknowledging the context in which a serious game will be played by a user and ensuring a relevant level of sensitivity to the distinct backgrounds, beliefs, and experiences of target users [21]. These are essential steps in justifying the use of a serious game, especially in the context of cognitive assessment.

The Cognitive Health of Older Adults in Prison

There has been a rapid increase in the number of people aging in prisons across the world in the last decade [22]. One area of particular concern for older adults in prison is age-related cognitive decline. There is evidence indicating that cognitive deficits are considerably underdetected in prison environments [23] and when they are detected, they are often managed informally or inadequately [24,25]. Studies on cognitive impairment in Australia, France, and the United Kingdom reported high levels of potential cognitive decline, with impairment rates of 15.3% (Australia) [26], 19% (France) [27], and 13% (the United Kingdom) [28] found in samples of people aged ≥50 years (≥45 years for Australian Aboriginal and Torres Strait Islander people) in prison.

Cognitive impairments lead to poor outcomes for older adults in the criminal justice system [23,29] and those transitioning out of prison [30]. Unidentified cognitive impairment can hamper a person’s ability to navigate the prison system and reduce their capacity to access (and routinely engage with) health and aged care services during incarceration and after being released from prison [31]. People with cognitive impairment are also prone to manipulation and coercion in prison, including physical and sexual trauma, peer pressure, and victimization [32]. People with cognitive impairments in prison can also have their symptoms perceived by staff and peers as purposefully obstructive or combative, resulting in punishment [33,34]. In addition, after being released from prison, people who have substantial cognitive impairments are less likely to adapt to life on the outside without appropriate support [24,25], resulting in increased risks of recidivism, homelessness, and hospitalization [30].

Although prison staff view early screening and diagnosis of dementia as a priority area for prison dementia care reforms [23,29,35], there still exists very limited cognitive assessment of older adults in prison, often because it is resource intensive. The use of a serious game for cognitive screening in prison appears to be a potential solution to the growing need for an effective and affordable cognitive assessment to support older adults in prison presenting with or at risk of cognitive deficits.

Conceptualizing Motivation to Play Games in Prison Using Self-determination Theory

If gamification promises to enhance user motivation to undertake cognitive assessment, but gamification does not always fulfill this promise, then a way to understand what increases (or decreases) a person’s motivation is necessary. A prominent theory used to understand a person’s motivation is self-determination theory (SDT). In SDT, the basic needs that ought to be satisfied to produce motivation, enjoyment, and...
overall well-being are competence (ie, experiencing mastery over challenges), autonomy (ie, doing something owing to an individual’s own volition), and relatedness (ie, experiencing meaningful social relations). In addition to helping researchers understand motivation generally, SDT has also been applied specifically to understand the motivation to play digital games [36-38]. The theory has been useful in explaining how game features relate to basic psychological needs and thus how game design and development decisions can increase (or decrease) user motivation and enjoyment [39]. Games have been shown to directly enhance motivation through features and gameplay [38].

In the video game context, competence refers to being challenged at the optimal level. For instance, a player will experience low competence if the challenges are very great (eg, game controls are difficult to use or enemies are very numerous) [37]. Autonomy may include the feeling of free choice to play a desired game and make choices within the game (eg, choosing which level to complete). Autonomy may also be satisfied when a user can design their own playable character and feel empowered to do things that they cannot do in real life. Relatedness can be achieved by playing with others, either in person or via the web [36]. Relatedness or a sense of belonging or connection can be satisfied by achieving team goals (eg, defeating the enemy together) or through healthy competition (eg, racing each other in Mario Kart).

SDT is a helpful framework to understand the motivation to play a serious game–based cognitive assessment in prison. Specifically, it provides a theoretical foundation to understand a specific user group’s fundamental psychological needs [40-42] and map these needs to tailored game design preferences [37,38]. In doing so, it ensures that a game is designed appropriately with the end user’s enjoyment, motivation, and well-being in mind [39]. Using SDT to query game design preferences and acceptability is an important exercise capable of expanding knowledge about desired user preferences for a serious game in prison.

Objectives of This Study

A critical first and ongoing step in developing suitable game-based assessment is to collaborate with the target population for whom the intervention is intended [43-45]. This point is especially important for older adults in prison whose preferences, values, and needs may deviate from those previously captured in studies on cognition and game design [46,47]. It is essential to capture and embed end-user input into game design and development to produce a game that is relevant to and compatible with the lives of older adults who are in prison.

Thus, the purpose of this qualitative thematic analysis was to understand the design preferences for a cognitive game-based assessment for older adults in prison, in the context of the prison environment and future transition back to community. Using the theoretical framework provided by SDT, the user design preferences for and user acceptability of a serious game–based assessment can be hopefully understood in a way that is useful for game designers to develop a cognitive assessment that is motivating, culturally suitable, nonthreatening, and enjoyable.

Methods

Philosophical Approach

This qualitative study was informed by a critical realist philosophy [48-51]. Critical realism broadly assumes that there are things that have a real, objective existence out there (ie, an independent world exists beyond our own constructions). However, critical realism appreciates that knowledge is fallible; our understanding of the world is based on and filtered through our own personalized interpretations [52]. Thus, critical realism makes an important distinction between epistemological and ontological assumptions. That is, ontology (ie, what is real and the nature of reality) is not reducible to epistemology (ie, our knowledge of reality) [48]. In other words, the critical realist approach combines realist ontology with constructivist epistemology [52]. After assuming this position, we can make ontological distinctions between our unique experiences and observations of events, the actual events themselves (which can be different from our experience of them), and the underlying mechanisms or structures that produce events [53]. Once we perceive reality in this way, we are open to exploring and explaining social events through reference to these mechanisms and structures and their effects.

Using the philosophical position of critical realism for this study has 2 major benefits. First, the deeply individual and subjective game preferences of older adults in prison can be interpreted empirically. Such an exercise can (and should) have direct input into game design considerations regarding the features, preferences, and esthetics of the game development process. Second, a more theoretical and critical analysis of participants’ observations and how the underlying mechanisms and structures may explain game preferences (and motivation for game-based cognitive assessments) can be undertaken. This second type of analysis is necessary to explore latent themes that may influence the success of a serious game–based intervention in the context of this complex population [54,55]. Considering both empirical observations and underlying mechanisms, simultaneously, and how they interact is critical to adequately design, develop, and deliver a game-based cognitive assessment for this marginalized population.

Sampling and Recruitment

Purposive sampling was used to produce representation across age, sex, and security level. Correctional staff identified and recruited potential participants after guidance from researchers to identify any older adults in prison, that is, aged ≥50 years (or ≥45 years for Aboriginal and Torres Strait Islander people). Aboriginal and Torres Strait Islander people are the Indigenous people of Australia, and relatively young ages are typically used in health, aged care, and policy settings to define older adults in this population, owing to population-level disparities in morbidity and mortality. Suitable participants were known to corrective staff through their day-to-day operations. Staff oversampled those who were potentially eligible for the study and shared the study information (study flyers and consent forms) with them. Consent forms were signed by participants in front of the prison staff. Potential participants were allowed to ask any clarifying questions to staff. After the consent forms
were signed, these were sent back to the research team for confirmation of informed consent.

We are sympathetic with the criticism by Braun and Clarke [56] about saturation as a useful concept in reflexive thematic analysis underpinned by constructionist epistemology. As such, we do not discuss our sample in terms of saturation. However, given our commitments to critical realist ontology, it is important to confirm the sufficiency of our sample in terms of information power [57] to support validity claims [52]. Thus, according to the criteria stipulated by Malterud et al [57], our sample size of 20 participants across 4 focus groups (FGs) was deemed sufficiently powerful given the (1) applied aim of the study, (2) sample specificity, (3) use of established theory (eg, SDT), (4) high quality of dialogue, and (5) comprehensive analysis strategy. As is standard, participants received AUD $15 (US $10.05) for lost work time. Funds were deposited into the inmates’ buy-up account.

Ethics Approval

Human research ethics approval was obtained from Corrective Services New South Wales Ethics Committee (D20/1014950), the University of New South Wales Human Research Ethics Committee (HC210546), Justice Health and Forensic Mental Health Network Human Research Ethics Committee (2021/ETH11114), and The Aboriginal Health and Medical Research Council (1873/21) before the commencement of the study.

Procedure

**FG Approach and Schedule**

We conducted 4 FGs, all of which were between 60 and 120 minutes long. All FGs were completed using internet-based audio-visual technology owing to the COVID-19 protocols. Sessions were led by the second author (AW), a neuropsychologist, and supported by RM and JH. FG questions were semistructured, and all 4 sessions followed a similar interview schedule. We were interested in game interventions but also more broadly in understanding people’s interaction with prison and past health services and their personal experiences with these services. After the objective of our study was discussed with participants, we showed a serious game demonstration (Graze Invaders; designed by AW, LM, and MK). This was used to provide a grounded example of what a serious game–based cognitive assessment might look like, in turn, setting some parameters for targeted design while still enabling participants to be open and creative [58] (refer to the study by Povey et al [21] for similar approach). After the general discussion about health and the demonstration, we asked targeted questions about what people thought about the serious game and then asked them to provide their general preferences for games and scrutinize the suitability (eg, logistically) of a game-based cognitive assessment in prison.

Example questions include the following:

1. Do you remember ever being asked questions about your memory or thinking while you have been in prison?
2. Have you ever had a memory or thinking test before coming to prison, maybe from a psychologist?
3. Before entering prison, did you ever use a computer or tablet?
4. Did you ever use a computer/phone/tablet to play games?
5. What kinds of extra things make you enjoy games more?
6. Do you think there would be any issues with you and other older adults in prison using this sort of brain game technology?
7. Is there anything else about using a serious game in prison you might like us to consider?

**Data Analysis**

This study used reflexive thematic analysis in which the active role of the researcher in producing knowledge is emphasized, with themes conceptualized as meaning-based patterns, rather than summaries of data [59]. Reflexive thematic analysis is compatible with our overarching philosophy of critical realism [60], which can be used to accurately explore the participant’s empirical world while engaging with underlying mechanisms and theories that can inform game design and development in the context of prison [44,61]. An abductive analytical approach was adopted to conduct this process. Abduction—also known as theoretical redescription—is a technique whereby empirical data are redescribed using theoretical concepts (in this case, SDT and game design research) [54,62]. Abduction has been defined as a process of “inference or thought operation, implying that a particular phenomenon or event is interpreted from a set of general ideas or concepts” [51]. Abduction raises the level of theoretical engagement beyond detailed description of the empirical entities but with an acknowledgment that the chosen theory is fallible [51]. This enables us to investigate participant observations in the context of latent, theoretical, and underlying mechanisms [51,53], which may be influencing game design preferences and motivations, while also ensuring that user observations and insights are still central to theme generation [56].

The analytical process followed the phases presented by Braun and Clarke [56,59,63]. Specifically, the first author transcribed and reread the interview transcripts and then coded and collated the interesting features of the data. Next, codes were added or updated deductively; that is, we explicitly analyzed our data through the lens of existing theory (namely, SDT). When we were satisfied with our coding, we used an abductive approach to combine and recontextualize existing codes. After coding and collating was completed, themes were drafted. These theory laden themes were analyzed to query any important underlying mechanisms, which may have been influencing the empirical observations. Finally, themes were organized and named in ways to appropriately portray the overall story of analysis. Using the process of abduction, all themes embody both the experiences of end users and relevant theories redescribing these observations in a context relevant to developing game design features [45].

**Validity**

While generating themes, the rigor and quality of the analysis can be enhanced by considering the types of validity (descriptive, interpretive, and theoretical) by Maxwell [52] and broad indicators of validity such as empirical adequacy, ontological plausibility, and explanatory power [64]. Interviews
were transcribed verbatim and then reread to check the accuracy of the transcription by the first author (RM), who also conducted the primary analysis. This addressed the questions regarding descriptive validity. The predetermined group design differences (women vs men; Aboriginal vs non-Aboriginal; and aged 50-64 years vs aged ≥65 years) in the FGs were used to establish consistency and key differences [49] in participant accounts from varied perspectives. The FG questions were partially informed by existing theoretical accounts on serious game design (eg, questions about game design features and typical duration of a game-based assessment); however, given the nature of FG sessions, general and open-ended questions were also asked, which enabled participants to discuss general ideas together and form intragroup consensus organically. The second author (AW) and senior author (JH) acted as critical friends [65] to the first author (RM), who provided critical dialogue and challenged the first author’s assumptions and explanations of phenomena at all stages of analysis (ie, initial coding, developing themes, and reviewing and naming themes). The use of the critical friend approach to thematic analysis underpinned by critical realism was conducted successfully by Goddard et al [44]. Finally, the applied basis for the research objectives (ie, to establish design preferences for an effective game-based assessment) produced clear practical utility [44,64] (this is discussed further in the Limitations section).

Results

Overview

Overall, 4 FGs with a total of 20 people were conducted with older adults (aged ≥50 years; aged ≥45 years for Aboriginal and Torres Strait Islander People) across 3 distinct prisons in New South Wales, Australia. Participants’ characteristics are presented in Table 1.

Overall, 3 distinct themes were generated. The first theme emphasizes the participants’ collective desire for an individually optimal level of difficulty in serious gameplay. The second theme raises the importance of avoiding immature and childlike gameplay features, as they can feel condescending to some older end users. The third theme highlights the strong user preference for meaningful choice and variety in any serious game–based cognitive assessment. According to SDT, the first theme relates to satisfying competence needs, whereas the second theme empathizes the need to avoid competence violations. The final theme is related to satisfying autonomy needs through meaningful choice.
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Group; security classification and participants</th>
<th>Age (years)</th>
<th>Time served</th>
<th>Identified as Aboriginal</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG1; maximum security</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>52</td>
<td>&lt;1 year</td>
<td>No</td>
<td>Female</td>
</tr>
<tr>
<td>P2</td>
<td>54</td>
<td>&lt;1 year</td>
<td>No</td>
<td>Female</td>
</tr>
<tr>
<td>P3</td>
<td>49</td>
<td>&lt;1 year</td>
<td>Yes</td>
<td>Female</td>
</tr>
<tr>
<td>P4</td>
<td>51</td>
<td>&lt;1 year</td>
<td>Yes</td>
<td>Female</td>
</tr>
<tr>
<td>P5</td>
<td>53</td>
<td>&lt;1 year</td>
<td>No</td>
<td>Female</td>
</tr>
<tr>
<td>P6</td>
<td>54</td>
<td>&lt;1 year</td>
<td>No</td>
<td>Female</td>
</tr>
<tr>
<td>FG2; maximum security</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>49</td>
<td>&gt;1 year</td>
<td>Yes</td>
<td>Male</td>
</tr>
<tr>
<td>P2</td>
<td>57</td>
<td>&gt;5 years</td>
<td>Yes</td>
<td>Male</td>
</tr>
<tr>
<td>P3</td>
<td>61</td>
<td>&gt;5 years</td>
<td>Yes</td>
<td>Male</td>
</tr>
<tr>
<td>P4</td>
<td>63</td>
<td>&gt;10 years</td>
<td>Yes</td>
<td>Male</td>
</tr>
<tr>
<td>FG3; maximum security</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>66</td>
<td>&gt;20 years</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P2</td>
<td>47</td>
<td>&gt;1 year</td>
<td>Yes</td>
<td>Male</td>
</tr>
<tr>
<td>P3</td>
<td>52</td>
<td>&gt;1 year</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P4</td>
<td>64</td>
<td>&gt;5 years</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P5</td>
<td>60</td>
<td>&gt;1 year</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>FG4; minimum/maximum security</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>72</td>
<td>&gt;1 year</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P2</td>
<td>76</td>
<td>&gt;5 years</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P3</td>
<td>82</td>
<td>&gt;1 year</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P4</td>
<td>90</td>
<td>&gt;5 years</td>
<td>No</td>
<td>Male</td>
</tr>
<tr>
<td>P5</td>
<td>69</td>
<td>&gt;5 years</td>
<td>No</td>
<td>Male</td>
</tr>
</tbody>
</table>

FG: focus group.

Goldilocks—Getting Gameplay Difficulty Just Right Through Optimal Challenge

There was considerable diversity in background, time served, education level, experience with computers and technology, and health status across the FG participants. This contributed to a variety of different expectations and preferences for what a serious game ought to be and the level of complexity that should be incorporated into gameplay.

One of the older men from FG4 felt that a simplified approach would be the most suitable:

I was in rehab 40 years ago...A couple of the blokes complained that it was a complicated program for simple people. And it should have been a simple program for complicated people. So, keep it simple stupid. [Participant 3; FG4]

In contrast, one of the women from FG1, who had plenty of previous experience with games, was interested in a challenging game-based assessment:

I’m a games freak! So, yeah, I play...Obviously it’s good if it gets harder, like to challenge us as you go, as you progress. Yeah, that would be good. [Participant 2; FG1]

This desire for a challenge through gameplay emerged several times across FG1, FG2, and FG3:

You need a challenge to keep people interested. [Participant 5; FG3]

However, FG4, which consisted of a group of older men (aged ≥65 years) residing in a Frail and Aged Unit, was keen to reiterate that whatever serious game was designed, it was best if it was “probably not too complicated” (participant 1; FG4).

One of the men in FG4 also added the following:

If you’re happy doing it, I don’t think it matters. [Participant 3; FG4]

This highlighted that if the level of complexity in gameplay was appropriate, participants would be motivated to play it.
This desire for an optimal difficulty level [37,38] reflected a broad consensus across all groups that if the game provided a suitable challenge, it would be far more enjoyable.

When discussing a game he enjoyed, a participant from FG2 reflected the following:

Because you need skills to do it, it is pretty good.
[Participant 2; FG2]

What constitutes a challenge is likely to be different for each end user. For instance, the need for a technical and skill-based challenge by participant 2 in FG2 was not the preference of another member of FG2:

The other ones are too hard...I stay with Solitaire.
[Participant 4; FG2]

With this desire for different levels of difficulty, participants agreed about the need for challenge flexibility to ensure that game difficulty was suitable for each player:

As long as it challenges you, I’d probably stay with it. [Participant 5; FG3]

...And if you can speed it up and slow it down, that gives people with no memory a better go at it. I can’t remember things very well, so that adjustment of the speed level is pretty good. That’s the only reason that I’d play that game. [Participant 4; FG3]

According to SDT, participant observations and consensus about the need for optimal challenge reflect the need to satisfy the basic psychological need for competence [37,38]. In the context of games, optimal challenges can lead to a sense of efficacy and reduces the likelihood of attentive engagement [39]. If a game is very easy, it becomes boring and reduces the likelihood of motivation, engagement, and enjoyment:

If you lose interest early, it is doomed. [Participant 5; FG3]

In contrast, if it is very hard, it can become frustrating and discouraging, thus also reducing the likelihood of motivation, engagement, and enjoyment:

My brain doesn’t work that quick. [Participant 4; FG2]

Optimal challenge is mediated by an individual’s threshold for difficulty [39,66]. Difficulty creates the challenge that is essential for satisfaction of competence needs. However, the right level of difficulty will be vastly different across individuals, especially in the diverse prison context:

Maybe you can design the project that you’re doing to different levels of how people want to react to these things. [Participant 1; FG3]

Given the participants’ need for optimal challenge, alongside the diversity in what challenge means for each person, a serious game–based assessment that can effectively differentiate game complexity appears to be important in the prison context to ensure that the challenge level is just right to address competence needs.

Avoiding Childish Graphics—Gimmicky Gameplay Can Be Condescending

Participants from FG2 and FG3 agreed that if a game did not have modern graphics or if it appeared childish, it was not playable and was even condescending:

No disrespect for who designed that game [the demo shown to participants] but that’s a game for a 9-to-10-year-old...no offense but I couldn’t play that...
[Participant 1; FG3]

I’m 60 years old and I use the phone and the graphics on the phone are better than that. You can’t expect people to go back, to be honest. [Participant 1; FG3]

It was clear that their expectations were for a game to feel more mature:

Your graphics are really bad [laughing]. Nintendo 64 was better than those graphics. [Participant 1; FG2]

[The game needs to be] more realistic. [Participant 5; FG3]

[And] more adult. [Participant 1; FG3]

In contrast, the women from FG1 had a far more positive response to the serious game demonstration, with only 5% (1/20) of the participants finding the game to be problematic:

It’s giving me anxiety. [Participant 3; FG1]

I like it. [Participant 4; FG1]

It’s cute. [Participant 2; FG1]

When queried about their general game preferences, the women in FG1 reacted positively to the ability to choose a character, even a character of the opposite sex. A participant suggested the following:

That’s half the fun in it!! [Participant 4; FG1]

The older men from the aged ≥65 years group were mostly indifferent to the game demonstration and did not comment about the graphics. They seemed to be more interested in functional issues connected to the game, such as whether it had audio and what the game was testing.

Although the openness to different game types varied across the groups, the strength of the aversion to childish games from several participants in FG2 and FG3 is important to be emphasized. Any game experience that makes a substantial portion of the target users feel condescended needs to be avoided:

I’d rather throw it out the window then just occupy my time with it. It has to be more than just the premise of being a game. In that regard, that’s condescending. I don’t consider myself unintelligent so, as someone said earlier, speak to me as an adult. Not as a ten-year-old kid. I understand there’s people who have learning problems and all the rest of it, but do I have to be spoken to in the same manner as them?
[Participant 1; FG3]
Games that felt childish and immature appeared to violate competence needs in FG2 and FG3 for some participants [37-39]. Although these negative responses were not evident in FG1 and FG4, the need to avoid competence violations is of high importance for any game-based assessment. This is especially true for a cognitive assessment game in the prison context, where people already report feeling overlooked and unheard [40-42].

Furthermore, if people are worried about their cognitive health before completing the game-based assessment, any risk of competence violations must be minimized, otherwise feelings of self-stigma or shame may be exacerbated [9-11]. A participant reflected the following:

*Just a thought. When you’re planning for this...this is where psychs get it wrong...Say what you mean, mean what you say. We’re grown men. Don’t dabble around. Get to your point. Get your yes or no. Factor that in when you’re planning this stuff. Don’t treat humans like children. Nothing worse. It’s patronising and condescending. Makes you feel stupider than you know you already are.* [Participant 2; FG3]

Thus, the need to ensure that game-based assessments in prison environments do not feel overly childlike or immature appears to have vital importance:

*As far as games and all that, we had the Atari and Mario Brothers and all that. We played it with the kids. It’s not just us playing. That was my involvement with games. But yeah, like the boys have said chess, backgammon, tiddlywinks, monopoly, we played all that. But I mean, games need to be more geared towards brain power, rather than just simplistic things like building blocks on top of other blocks like Tetris. But yeah, I got no problems with them.* [Participant 1; FG3]

Game variety appeared to be a key preference for participants:

*It doesn’t matter who you are, you can see on the screen, ok, I don’t know how to do that, but I’d like to know, so I’ll go and play that one...or, ok...I know how to do that, so I’ll go play that other one.* [Participant 1; FG2]

*Yeah, you got to have a variety.* [Participant 4; FG2]

Participants were also quick to acknowledge and discuss the purpose of the game, that is, to test different types of cognitive functioning (reaction time, memory, executive ability, etc). By acknowledging the primary purpose of the game, participants saw the opportunity to create a variety of gamified tasks addressing different design appetites within the groups and within themselves:

*Yeah, you could have mind games, where you gotta think about what you’re doing. And reactive games, where you gotta react to something and all that sort of thing...you could have literacy sort of games where you gotta think about what’s being written and other games where you’re using your reaction time and you’re thinking about what’s going on but you’re not reading it.* [Participant 1; FG2]

When brainstorming about game design preferences, a participant from FG2 even suggested the technique of game customization to avoid amotivation and maximize meaningful choice:

*Ohh, I just come up with a thing you can do for what you need. If you got someone who is getting dementia and they are retired, and you knew what their trade was, with the trade they’d done for work, you could ask them things like...let’s say they done carpentry...Say, “what kind of tools would I need to build this kind of table and the process that I need to go through to start [to] build that table?” And you could see what their memory is from what they used to do. If you use to do something, your memory should recall, and you ask them something about that.* [Participant 2; FG2]

Another member of FG2 expanded upon this customization idea:

*Going back to the building side of things, you could have a Mah-jong sort of game with building materials. Even engineering, you could do it with cars, make it like you gotta build a car or something.* [Participant 1; FG2]

Throughout the brainstorming conducted across a number of FGs, especially FG2, the game design technique of minigames emerged as a viable solution to address the need for variety and meaningful choice across the serious game.

When asked about the potential to use minigames, a participant positively reflected the following:

*Like little minigames where, between what’s going on around them, they’re coming down like a map of different quests type thing? You’re progressing through, or something, and play a little mini...*
This preference for different minigames reinforces the appetite for meaningful choice, a critical element of autonomy in the context of game design [37-39]. According to SDT, autonomy in digital games involves interesting options and volitional engagement [67]. Although choice and variety are critical design elements in most games, the strong desire for these elements by participants is important to be highlighted in the context of prison—an environment where autonomy is often thwarted [41].

Discussion

Principal Findings

This study aimed to qualitatively examine the design preferences for a serious game–based cognitive assessment among older adults in prison. Using an inductive approach, we generated 3 distinct themes. They were (1) Goldilocks—getting gameplay difficulty just right through optimal challenge, (2) avoiding childish graphics—gimmicky gameplay can be condescending, and (3) a balanced diet—meaningful choice and variety keeps game-based assessments fun. These themes provide novel insights into the game design preferences of older adults in prison, whose voices have been seldom heard in the development of serious games so far. In addition, the production of these themes in the context of SDT and game design research should, we hope, provide a useful and actionable base for serious game designers to develop appropriate cognitive games that are acceptable and motivating for older adults within the unique prison context.

Within theme 1, it was clear that the competence needs for an individually optimal level of difficulty was a priority across all FGs. Regarding game design, this can be addressed by developers through dynamic game balancing [66,68]. Dynamic game balancing is a popular game design technique that overcomes variability in individual difficulty thresholds through real-time adjustment of game parameters, so that task difficulty adapts and eventually aligns with a player’s ability [66]. This ensures that competence can be satisfied for a variety of users through optimal challenge [69]. The goal is to keep the user interested from beginning to the end. Dynamic game balancing is associated with high levels of user satisfaction [70] and feelings of competence, which predicates high levels of motivation and enjoyment [69]. The use of dynamic game balancing for serious game–based cognitive assessments is an emerging area, and potential complexities such as maintaining psychometric validity are yet to be fully considered [66]. Nonetheless, the technique appears to be theoretically capable of addressing the competence needs for optimal challenge highlighted by participants.

Theme 2 reflected the need to avoid competence violations by minimizing childlike game features, which made some participants feel condescended. In the context of SDT, game features offering opportunities for mastery that provide optimal challenges can satisfy competence needs by helping a player feel a sense of accomplishment and control [38]. However, the opposite can also be true. When people are tasked with playing a game that they believe is not suitable for them or does not align with their self-perceived level of capability, it can become boring, frustrating, and even insulting. This is an especially important consideration in the context of prison, where the need for competence is often thwarted, and people often distrust the prison system and report feeling disrespected by health and corrective professionals [71]. A noteworthy minority of our study participants highlighted that childish graphics made them feel condescended, and some members of our FGs appeared to be frustrated by the idea of playing a childish game for the purpose of cognitive assessment. We suggest that this frustration stems, at least in part, from a misalignment between some participants’ self-perceived level of capability and the perceived inability of a game that looks and feels childish to meet this capability. We further suggest that this frustration, a common and natural response to perceiving a game as incompatible with one’s intellectual capacity, can be exacerbated by the competence-thwarting prison environment, in which control is already difficult to achieve [72]. Thus, avoiding games that feel similar to kids’ games and designing features that are more adult and realistic, which are geared toward brain power appear to be important to avoid competence violations and, more broadly, guard against overly negative responses to serious game–based assessments in prison.

Theme 3 highlighted the strong user preference for choice and variety, both of which satisfy the need for autonomy and can enhance intrinsic motivation [37,38,67]. Although the positive effects of choice can be complex and depend on a variety of factors (eg, cultural background), multiple studies have shown that choice has a beneficial effect on motivational outcomes such as effort and task performance [41,73]. Autonomy satisfaction has also been associated with better quality of life in prison [41]. However, both autonomy [74,75] and choice [76] are often difficult to attain in prison [41]. Therefore, ensuring that games are autonomy-focused through meaningful choice when they are restricted in real life may be particularly beneficial for this group. Minigames are an effective game technique to maximize choice and variety [38,39] and, as highlighted by participants, align well with the objectives of cognitive assessment. Thus, designing minigames with emphasis on user control appears to be a suitable way to maximize in-game autonomy for people from marginalized groups who may struggle to achieve autonomy in other aspects of their lives.

Limitations

This study has some important limitations. First, this study purposefully sampled end users from 3 different prisons in New South Wales, Australia. Although attempts were made to select a distinct variety of older adults in prison and in turn understand empirical observations in terms of prominent theories or underlying structures, the themes generated are not necessarily generalizable to other locations or user groups. Second, all FGs were monitored by a correctional officer owing to security protocols. This power imbalance may have affected the responses given by participants, particularly those who are highly reliant on staff, such as those in the Frail and Aged Unit. However, no evidence of this was observed during the FGs. Third, the method of analysis used in this study and our interpretation of data are underpinned by the philosophy of
critical realism. It is entirely possible that a researcher informed by a different philosophical approach would interpret the empirical observations differently and thus make different truth claims. Similarly, results are reported using the technique of abduction. We have made attempts to be clear and transparent about our theory-focused interpretations and recontextualization of the data when producing our results. However, it is critical to acknowledge that our interpretations are inevitably based on our own beliefs and objectives. In addition, our chosen theories are fallible. To overcome this, we attempted to reduce biases and improve validity (refer to the Validity section); however, these attempts are not perfect. Moreover, they are not intended to remove the researcher completely from the qualitative process. With this in mind, it should be noted that the authors of this paper are focused on building a serious game for cognitive assessment that is suitable for use by diverse older adults in prison. We want the game to be successful and have likely interpreted participant observations and preferences to ask how to make a serious game work rather than whether serious games work? This does not mean we are certain that a game will be valid, feasible, or acceptable in prison. However, we certainly hope it will be, and this desire to design and develop a suitable serious game has informed this study.

Practical and Research Implications

A serious game–based assessment provides a novel challenge for game designers and developers. This is because the intended application of a serious game assessment is likely to be a relatively short (eg, 10-20 minutes) user experience, which is only played by users sporadically. The implication is that traditional game design focusing on intrinsically motivating features such as narrative and social connectedness may not be feasible or desirable. For instance, a complex and value-based storyline, with unique characters and quests, is likely very complex for an assessment that someone is going to undertake irregularly for 10 to 20 minutes. This tension has not been given much attention in the serious game literature from a user design perspective. Most of the previous studies on user preferences of cognitive game design have highlighted the importance of embedding intrinsically motivating game features without considering the complexity of embedding these features into a very short and sporadically played game-based assessment [77]. These traditional features may be desirable in long-form games focused on cognitive training or rehabilitation but are less suitable for the narrow and short objectives of a cognitive assessment. In contrast, other attempts to make game-based cognitive assessments have gamified traditional tasks in very basic ways. Although this may be more suitable to ensure psychometric validity, this process risks not being particularly engaging or fun. For instance, some previous attempts to digitize cognitive tasks or introduce simple gamification have been unsuccessful, with users reporting that, similar to traditional tasks, the experiences were tedious and boring [3,20,78,79].

Our study was guided in part by the need to address this tension. We were open with our participants about the typical length of serious games and the intended purpose of a gamified cognitive assessment (ie, to detect cognitive decline or dementia). This enabled users to provide practical design insights that reflect the unique parameters of serious game–based cognitive assessment. As such, the themes generated are consistent with the reality of serious game–based design and conform to the nuances of gamifying short and sporadic assessments while still highlighting game design features that are immersive, motivating, and enjoyable to our distinct user group. In this way, optimizing the challenge by balancing the difficulty level, avoiding childish features, and maximizing meaningful choice through minigames are unique user-generated design preferences suited to addressing user needs within a short and potentially infrequent serious game–based assessment in the prison setting.

Conflicts of Interest

MK is the chief executive officer of Arludo.

References


2. 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 Sep 08;8(3):e17810 [FREE Full text] [doi: 10.2196/17810] [Medline: 32897233]


Abbreviations

FG: focus group

SDT: self-determination theory
Abstract

Background: Cognitive assessment using tangible objects can measure fine motor and hand-eye coordination skills along with other cognitive domains. Administering such tests is often expensive, labor-intensive, and error prone owing to manual recording and potential subjectivity. Automating the administration and scoring processes can address these difficulties while reducing time and cost. e-Cube is a new vision-based, computerized cognitive assessment tool that integrates computational measures of play complexity and item generators to enable automated and adaptive testing. The e-Cube games use a set of cubes, and the system tracks the movements and locations of these cubes as manipulated by the player.

Objective: The primary objectives of the study were to validate the play complexity measures that form the basis of developing the adaptive assessment system and evaluate the preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool.

Methods: This study used 6 e-Cube games, namely, Assembly, Shape-Matching, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze, each targeting different cognitive domains. In total, 2 versions of the games, the fixed version with predetermined sets of items and the adaptive version using the autonomous item generators, were prepared for comparative evaluation. Enrolled participants (N=80; aged 18-60 years) were divided into 2 groups: 48% (38/80) of the participants in the fixed group and 52% (42/80) in the adaptive group. Each was administered the 6 e-Cube games; 3 subtests of the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; Block Design, Digit Span, and Matrix Reasoning); and the System Usability Scale (SUS). Statistical analyses at the 95% significance level were applied.

Results: The play complexity values were correlated with the performance indicators (ie, correctness and completion time). The adaptive e-Cube games were correlated with the WAIS-IV subtests (r=0.49, 95% CI 0.21-0.70; P<.001 for Assembly and Block Design; r=0.34, 95% CI 0.03-0.59; P=.03 for Shape-Matching and Matrix Reasoning; r=0.51, 95% CI 0.24-0.72; P<.001 for Spatial-Memory and Digit Span; r=0.45, 95% CI 0.16-0.67; P=.003 for Path-Tracking and Block Design; and r=0.45, 95% CI 0.16-0.67; P=.003 for Path-Tracking and Matrix Reasoning). The fixed version showed weaker correlations with the WAIS-IV subtests. The e-Cube system showed a low false detection rate (6/5990, 0.1%) and was determined to be usable, with an average SUS score of 86.01 (SD 8.75).

Conclusions: The correlations between the play complexity values and performance indicators supported the validity of the play complexity measures. Correlations between the adaptive e-Cube games and the WAIS-IV subtests demonstrated the potential utility of the e-Cube games for cognitive assessment, but a further validation study is needed to confirm this. The low false detection rate and high SUS scores indicated that e-Cube is technically reliable and usable.
cognitive assessment; computerized block games; adaptive serious games; computerized cognitive assessment

Introduction

Background

Cognitive assessment aims to measure multiple domains of cognition, including visuospatial abilities, working memory, language, attention, executive function, fine motor skills, and orientation [1]. One’s cognitive abilities affect learning outcomes, physical and mental health, social behavior, and interaction with the environment [2-4]. Identifying impairment in any of these domains, diagnosing the cause, specifying the severity, and tracking the progression of the symptoms are the common purposes of cognitive assessment in clinical settings [5]. This paper presents an innovative technology called e-Cube for adaptive, automated cognitive testing and reports the evaluation results in terms of preliminary utility and usability.

There are standardized instruments widely used for cognitive assessment. The Wechsler Adult Intelligence Scale (WAIS) has been broadly adopted in clinical, research, and educational settings and is often referred to as a gold standard [6]. The WAIS Fourth Edition (WAIS-IV) is normed for the ages of 16 to 90 years. It comprehensively assesses cognitive abilities using 15 subtests that target various cognitive domains [7]. This instrument is administered and scored by a qualified psychologist, taking approximately 60 to 90 minutes. This process is labor-intensive and costly [8]. The Stanford-Binet Intelligence Scales, Fifth Edition, is another standardized instrument commonly used in both clinical and research settings [9]. Several WAIS and Stanford-Binet Intelligence Scales subtests rely heavily on a person’s verbal skills and, therefore, show limitations when administered using a non–native language version [10]. There are also nonverbal instruments aiming to eliminate cultural and language biases in the assessment. For example, the Raven Progressive Matrices consist of 60 items measuring the basic cognitive functioning of individuals, each of which is a visual geometric design with a missing piece [11,12].

The advancements in digital technologies have enabled researchers to explore computer-based methods for cognitive assessment. Computer-based methods can reduce the administrative burden, automate the scoring process, reduce cheating, and standardize test conditions once successfully validated [13]. A straightforward application is to convert a paper-and-pencil test into a computerized version while retaining the contents and formats. Q-interactive is a digital system initially developed for the WAIS-IV that uses 2 iPads, one for the administrator and the other for the test taker [14]. This digital version reduces labor-intensity but takes approximately the same time for a trained professional. Moreover, it can only automate some types of tests. In particular, one of the subtests, Block Design (BD), requires the examinee to assemble physical blocks to match the top surface with a given image displayed on an iPad. The administrator then has to check the correctness and input the results manually. In addition to the computerization of existing instruments, an increasing body of research has adopted the concept of computer- or tablet-based serious games to make the experience more engaging [15-17]. Some serious games use dynamic difficulty adjustment to achieve adaptive testing by tuning item difficulty autonomously [18,19]. However, most of the previously developed games for cognitive assessment do not include measurements of fine motor and hand-eye coordination skills.

Cognitive Assessment Using Tangible Objects

Cognitive assessment sometimes uses tangible objects to measure one’s cognitive skills together with fine motor and hand-eye coordination skills. These skills are closely linked to many neurological diseases and brain injuries [20]. Existing research also suggests that the deterioration of fine motor control and coordination characterizes sensorimotor deficiencies in mild cognitive impairment and Alzheimer disease [15,21-23]. The BD subtest in the WAIS [5] and the Kohs Block Design test [24] use a set of cubes and require an examinee to place and assemble the top surfaces of the blocks to match the given image. Unlike the simple multiple-choice questions used in many other assessment instruments, the administrator has to inspect the correctness of the block manipulation visually while timing in these tests. This is labor-intensive and error prone owing to manual recording and subjectivity, possibly affecting the assessment results.

Automating the assessment using physical objects also involves additional challenges and requires technological innovations beyond what is expected for computerized tests. For example, a platform called ETAN supports the use of tangible user interfaces and physical objects for evaluating visuospatial cognition by implementing the Baking Tray Task [25]. Cognitive Cubes were designed to assess spatial and constructive abilities by asking users to build 3D shapes with the cubes. A pilot study involving 16 participants demonstrated that the Cognitive Cubes were sensitive to differences in cognitive ability [20].

SIG-Blocks and TAG-Game, developed for the automated assessment of cognitive and fine motor skills, were the previous research of this work [26-28]. Each SIG-Block, covered with simple black-and-white geometric shapes, can sense physical motions applied to it, detect adjacent blocks, and send sensor data to a local host computer in real time. TAG-Games are computerized games that use SIG-Blocks as a means of game control. In total, 3 types of TAG-Games, namely, Assembly, Shape-Matching, and Memory, were designed and tested. These games are all nonverbal and require hand manipulation of physical blocks. The TAG-Game technology is one of the few systems that can automate the administration and data collection of tasks involving physical object manipulation. However, despite its demonstrated potential, several challenges were identified in our previous research. Specifically, hardware costs, occasional technical failure, and high maintenance make the system unsuitable for broad and long-term adoption and use.
e-Cube Games for Automated Assessment of Fine Motor and Cognitive Skills

e-Cube is our latest technical innovation that converts the original TAG-Game system into a computer vision-based system using a set of plastic cubes and a webcam. The e-Cube system reduces the device cost from US $1500 to approximately US $50 (excluding a computing device needed for any computerized assessment), decreases potential technical errors, and nearly eliminates the maintenance burden. The entire system is fully autonomous and easy to use. In addition to these benefits, a new adaptive test environment was established based on the embedded algorithms for measuring play complexity and generating adaptive test items autonomously. These features enable personalized assessment based on an individual’s real-time performance. e-Cube consists of 6 types of games: (1) Assembly, (2) Shape-Matching, (3) Sequence-Memory, (4) Spatial-Memory, (5) Path-Tracking, and (6) Maze. The first 3 were directly adopted and converted from TAG-Games, and the other 3 were newly created. New computational measures of play complexity were defined and implemented for each game.

The evaluation focused on testing 2 objectives. Objective 1 was to validate the proposed play complexity measures that form the algorithmic basis of the adaptive games. Correlation analyses were performed between the developed play complexity measures and 2 performance indicators, mean correctness and mean completion time. Objective 2 was to understand the preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool. The non–age-corrected raw scores of 3 WAIS-IV subtests—BD, Digit Span (DS), and Matrix Reasoning (MR)—were adopted to compare their results with the e-Cube game scores. The WAIS-IV is a well-established instrument, and the 3 selected subtests measure the target cognitive domains of the e-Cube games. Specifically, the Assembly game was conjectured to be related to BD as both require the assembly of block surfaces to match a given pattern. The Shape-Matching game requires the participant to find a shape that completes a pattern, so it was expected to tap the same cognitive abilities as the MR subtest. Sequence-Memory and Spatial-Memory were expected to be related to DS as they all target working memory skills. The remaining games, Path-Tracking and Maze, are timed games asking participants to give the shortest trajectory by reasoning, so they were both hypothesized to show a relationship with BD and MR. The hypothesized relationships between the e-Cube games and the WAIS subtests are summarized in Table 1. The false detection rate of the system determines whether it produces reliable and accurate data. Usability was evaluated by administering the System Usability Scale (SUS) to all participants upon the completion of the assessment session. The SUS is a 10-item questionnaire measuring usability with high validity and reliability and, thus, used as a measure of perceived usability [29-31].

Table 1. The 6 e-Cube games with their associated task descriptions and the expected associations with the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), subtests (Block Design [BD], Digit Span [DS], and Matrix Reasoning [MR]).

<table>
<thead>
<tr>
<th>e-Cube game</th>
<th>Task</th>
<th>WAIS-IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly</td>
<td>Assemble multiple cubes to match the top assembly configuration</td>
<td>✓</td>
</tr>
<tr>
<td>Shape-Matching</td>
<td>Manipulate 1 cube to complete the pattern with 1 missing piece</td>
<td>✓</td>
</tr>
<tr>
<td>Sequence-Memory</td>
<td>Memorize a sequence of geometric shapes and reconstruct it using 1 cube</td>
<td>✓</td>
</tr>
<tr>
<td>Spatial-Memory</td>
<td>Memorize a spatial assembly of geometric shapes and reconstruct it using cubes</td>
<td>✓</td>
</tr>
<tr>
<td>Path-Tracking</td>
<td>Trace a connected path between 2 blue dots using a single cube</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>Maze</td>
<td>Navigate through a maze to reach a goal point from a starting point using a single cube</td>
<td>✓ ✓</td>
</tr>
</tbody>
</table>

Methods

e-Cube Games

System Overview

The e-Cube system consists of a set of 9 cubes with 1.2-inch–length edges, a place mat with a brown rectangular region in the center, a computing device with a display, and a webcam with a custom-designed stand (Figure 1). The cube’s 6 faces are distinctive black-and-white geometric shapes, including squares, strips, and triangles representing 4-, 2-, and 1-fold rotational symmetry (Figure 2). The cubes preserve the same design as the SIG-Blocks [28]. When the system turns on, the camera automatically detects the corners of the brown rectangular area on the place mat. This area is called the play area, where the cubes are expected to be placed and manipulated. The laptop with the connected webcam displays the cubes in the play area after perspective transformation projecting the original camera view onto a 2D plane and tracks their movements in real time [32]. This autonomous transformation offers flexibility in the camera location.

The e-Cube system requires the accurate identification of the top-surface images of the cubes. Individual box-shaped regions are assigned for placing the cubes (Figure 1), wherein the geometric shape and orientation detection algorithm is executed. The embedded algorithm first detects the black-and-white regions within each box to check whether a cube exists. It then identifies a polygon using the Ramer-Douglas-Peucker algorithm [33]. Finally, the specific shape and orientation of the detected polygon are determined. This simple strategy makes the system robust and reliable under different illumination conditions and limited computing capabilities. The e-Cube system can run reliably on a relatively low-end computing device, such as Intel Core i5-7200U (2.5 GHz, 3 M cache, dual core, and 4 threads).
Figure 1. The hardware of e-Cube, consisting of the cubes, a webcam with a stand, a place mat, and a host computing device running the Assembly game.

![Image of e-Cube hardware](image)

Figure 2. A total of 9 geometric cubes and 14 distinctive surface shapes with their IDs formed by rotating the images on the 6 surfaces of a cube by 0°, 90°, 180°, and 270°.

![Image of geometric cubes and surface shapes](image)

<table>
<thead>
<tr>
<th>Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDs</td>
<td>(0,0,0)</td>
<td>(0,1,0)</td>
<td>(1,0,0)</td>
<td>(1,1,0)</td>
<td>(1,1,1)</td>
<td>(2,0,0)</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 presents the tasks associated with each game, and Figure 3 shows an example item for each e-Cube game. Assembly asks the player to match the given assembly figure displayed on the screen using 4 or 9 cubes, similar to the BD subtest of the WAIS-IV. Shape-Matching involves items with assembly patterns, each missing 1 piece, and the player completes the pattern using a single cube. Sequence-Memory and Spatial-Memory require the player to memorize a sequence or an assembly of geometric shapes. In Sequence-Memory, each shape is displayed for 1 second and then disappears. In Spatial-Memory, an assembly pattern of 2, 3, or 4 geometric shapes is displayed for 5 seconds. The items in the Spatial-Memory game are similar to those in Assembly, whereas visible outlines around individual shapes are added, as shown in Figure 3, to assist perceptual segmentation of the pattern [34]. Path-Tracking and Maze use only 1 cube with its white square facing up. In these 2 games, the vision algorithm detects the center of the white square and tracks it continuously on the screen; no assigned box-shaped regions are shown on the screen. Path-Tracking displays a green connected path between 2 blue dots, and the player must trace the path by moving the cube from one blue dot to the other on a 5 x 5 grid via the shortest path. The Maze game asks the player to find the shortest path of mazes shown on the screen by moving a cube from the start (blue) to the end (red).

Figure 3. Sample items for the 6 games.

![Sample items for each game](image)
**Computational Measures of Play Complexity**

**Overview**

The e-Cube system aims to dynamically adapt to individual differences in cognitive skills by generating test items autonomously based on one’s real-time performance. To do so, a computational method to measure the difficulty of each item is required. The previously defined measures of play complexity presented in the studies by Lee et al [26,35] and Jeong et al [28] were highly correlated with the participants’ performances measured using completion time or accuracy. These measures captured the complexities associated with individual geometric shapes without considering the spatial complexity of the assembly patterns. For example, the 3 assembly patterns shown in Figure 4 had the same complexity value using the previously defined measures. As our previous study used a handcrafted set of items, we could select the items where their difficulties could be clearly differentiated using the previously defined measures. However, for generating adaptive test items, the complexity measures must capture the difficulties associated with both the individual shapes and the assembly patterns.

**Figure 4.** Items formed by the same geometric shapes but with different play complexity (with identical compositional complexity but different configurational complexity).

To address this limitation, we defined new complexity measures for the 6 e-Cube games. Two mathematical concepts were applied: (1) the Shannon entropy and (2) the gray-level co-occurrence matrix (GLCM). The Shannon entropy measures the uncertainty, randomness, or disorder existing in the data [36] and is calculated as

$$H = -\sum_{i} p_i \log p_i$$

where $p_i$ is the probability of the $i$th event. When the probabilities are evenly distributed, the Shannon entropy is calculated as $H = \log n$. The GLCM was originally proposed to classify image texture in grayscale [37,38]. For an image with an $m \times n$ dimension and $L$ gray level, the GLCM of the image ($f$) is defined as an $L \times L$ square matrix such that

$$[f(x,y)]_{ij} = \frac{Q}{m \times n}$$

where $\Delta x$ and $\Delta y$ are typically defined as the horizontal, vertical, or diagonal position differences between the 2 adjacent pixels [37]. Horizontally adjacent pixels can be paired along 0° or 180°; vertically adjacent pixels can be paired along 90° or 270°; and diagonally adjacent pixels can be paired along 45°, 135°, 225°, or 315°. On the basis of the Shannon entropy and the GLCM, the computational measures of play complexity for the 6 e-Cube games are defined in the following sections.

**Play Complexities of Assembly, Sequence-Memory, Spatial-Memory, and Shape-Matching**

The play complexity of the items in Assembly, Sequence-Memory, and Spatial-Memory is computed using

$$C_{\text{compos}}$$

where $C_{\text{compos}}$ represents the compositional complexity associated with individual shapes (ie, the number of shapes and their rotational symmetry). $C_{\text{config}}$ captures the configurational complexity associated with the orientation and color differences among the shapes in the way that they are arranged, and $k$ is a sigmoid function defined as

$$k(x) = \frac{1}{1 + e^{-x}}$$

If $C_{\text{config}}$ is small, a small $k$ leads to a lower impact of $C_{\text{compos}}$ on $C_{\text{play}}$. For example, if an item is formed only by identical triangles (large $C_{\text{compos}}$ and small $C_{\text{config}}$), $C_{\text{play}}$ will still be small owing to $k$.

The Shannon entropy forms the basis of $C_{\text{compos}}$ such that

$$C_{\text{compos}} = \sum_{i} p_i \log p_i$$

where $Q$ is the total number of shapes in the item, $m_i$ is the number of available distinctive shapes among the 6 faces of a cube ($m_i=6$ if all faces of a cube are different), and $r_i$ is the number of distinctive orientations obtained by rotating this shape 90° ($r_i=1$ for squares, 2 for strips, or 4 for triangles). The 3 images in Figure 4 have the same $C_{\text{compos}}$ value.

The GLCM was adopted for capturing the configurational disorder ($C_{\text{config}}$) [39]. Figure 5 illustrates how it is obtained for an Assembly item. First, all geometric shapes used in each item are represented as $J$ with the indexes corresponding to each shape defined in Figure 2 and its location. Second, all adjacent pairs along the 0°, 45°, 90°, and 135° directions on $J$ are extracted. For example, the second row in $J$ is (3, 5, 4), and the ordered pairs along 0°, including the circumulant pair, are (3, 5), (5, 4), and (4, 3). Once all pairs are obtained, the number of each pair is imposed on the location in a 14-by-14 matrix $f$, that is, the GLCM, following equation 1. As shown in Figure 5, there are two (1, 5) pairs that correspond to 2 in the (1, 5) coordinate in $f$ and one (1, 1) pair that corresponds to 1 in the (1, 1) coordinate. The weighted entropy [40] based on $f$ is then calculated using

$$E_{\text{weighted}} = \sum_{i} w_i f_i$$

where

$$w_i = \frac{1}{r_i}$$
The weight $w_{ij}$ estimates the configurational complexity of 2 adjacent elements based on their colors and orientations. To compare the differences among these 14 distinctive shapes, 3 IDs were assigned to each shape to categorize its geometric shape (square, strip, or triangle), color, and orientation (Figure 2). Regarding the IDs, 2 was assigned to the weight if the 2 adjacent shapes had different colors and orientations, and 1 was assigned otherwise.

In Shape-Matching, as the player was asked to find a single shape that best completed the pattern, more shapes used in the pattern do not necessarily indicate greater difficulty in the pattern. Therefore, we only used the configurational complexity to estimate the item difficulty such that $C_{\text{config}}$ is defined as the summation of the weighted entropies based on the 3 GLCMs estimating how frequently a pair occurs horizontally, vertically, and diagonally.

**Play Complexities of Path-Tracking and Maze**

Path-Tracking and Maze do not use the geometric shapes of the cubes and, instead, use a single cube for creating a path. Therefore, the aforementioned method is not applicable. The play complexity of Path-Tracking adopts the network complexity based on the Shannon entropy [41], given by

$$C_l = \sum_i s_i \cdot \log_2 s_i$$

where $V$ is the number of vertices and $a_i$ is the associated vertex degree. For Maze, the play complexity is defined as

$$C_s = \sum_i s_i \cdot \log_2 s_i$$

where $C_m$ is the maze complexity using equation 5 and $C_s$ and $C_l$ are calculated using

$$C_m = \text{length of the shortest path}$$

The computational measures of play complexity form the basis of the adaptive algorithms, which can automatically generate test items. On the basis of the concept of dynamic difficulty adjustment, we created an adaptive e-Cube system that can adjust the item difficulty based on a player’s performance measured using correctness.

The game begins with an item with a predefined low complexity. If the player answers the first item correctly, it proceeds to the next item with a higher complexity; otherwise, a new item with the same complexity is generated. If 2 consecutive incorrect answers are received, the complexity reverts to the midpoint between the latest correctly answered item complexity and the current incorrectly answered item complexity. The difference between the current and the next complexity value is referred to as a step size that can be either positive, 0, or negative. The game ends at a predefined highest complexity level or when the absolute value of the step size becomes sufficiently small.

The item generators for all games except for Shape-Matching follow a similar process, shown in Figure 6. The system takes a desired play complexity value $C_d$ and a small tolerance $e$ as input and generates a new item with a complexity $C_{\text{play}}$, where $|C_d - C_{\text{play}}| \leq e$. Apart from $C_d$ and $e$, additional input is needed in these games except for Maze. This input is the dimension of the pattern (e.g., 2 × 2 or 3 × 3) in Assembly and Spatial-Memory, the number of images to be displayed in Sequence-Memory, and the number of dots to be connected (referred to as nodes) in Path-Tracking. The following steps generate items: (1) the system randomly generates an item based on the inputs; (2) the absolute difference between $C_d$ and $C_{\text{play}}$ is computed, where $C_{\text{play}}$ is the complexity of the current item computed using the proposed measure; (3a) if the absolute difference is smaller than $e$, the system outputs the current item and ends the process; and (3b) if the absolute difference is not smaller than $e$, the system updates one feature of the current item to make the item easier or harder and then goes back to step 2. The features of the item can be geometric shapes in Assembly, Sequence-Memory, and Spatial-Memory; the paths connected by nodes in Path-Tracking; or the position of the end point in Maze.
Shape-Matching uses assembly configurations with embedded patterns where the types of patterns are predefined in the item generator, such as symmetry and rotation. Shape-Matching generates items from a predefined pool. For example, the easiest pattern in the predefined pool is formed by 4 identical shapes, in which one of the shapes will be hidden from players and treated as the missing piece. The item generator for Shape-Matching randomly selects a shape to form the easiest pattern, which leads to different items with the same play complexity.

Evaluation of e-Cube

The evaluation study focused on the preliminary validation of (1) the proposed play complexity measures that form a basis for developing adaptive games (objective 1) and (2) the preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool (objective 2).

Materials and Methods

The study used 2 versions of e-Cube games: e-Cube with the item generators (called adaptive e-Cube) and e-Cube with fixed items (called fixed e-Cube). Each participant was assigned to 1 of the 2 groups to experience the adaptive or fixed e-Cube games (ie, adaptive group and fixed group). The fixed games provided the same items for each player, whereas the adaptive games offered different items and different numbers of items based on the players’ performance. The fixed versions of Assembly and Shape-Matching used the same items as in the study by Lee et al [26]. The fixed items of the rest of the games are shown in Figure 7.

Objective 1 was tested by performing correlation analyses between the play complexity measures and performance indicators, including the mean correctness and mean completion time obtained by the participants for individual items in the fixed group. For objective 2, the correlations between the raw scores of 3 WAIS-IV subtests (BD, DS, and MR) and the 6 e-Cube games were analyzed to understand their relationships.

We also investigated the technical reliability of the system using the false detection rate and usability based on the SUS results.

Protocol and Recruitment

This human participant study took place at Texas A&M University (TAMU). Bulk recruitment emails were sent to TAMU communities, and flyers were placed in buildings within...
the university for recruiting healthy participants aged 18 to 64 years. Once potential participants contacted the research team, a prescreening survey was sent via email to self-identify their eligibility before scheduling a visit. The prescreening survey consisted of 4 questions on age, date of birth, sex, and health conditions. Individuals who were beyond the target age range or had any of the following health conditions were excluded: stroke, other neurological diseases, low vision or blindness with aid, hearing loss or deafness with aid, or difficulties in arm or hand movements for manipulating small objects.

The sample size of a main trial is usually determined through a power analysis, where the variance is known from previous or pilot studies [43]. However, for this preliminary study, we applied the simplest method—sample size rules of thumb, which recommended samples of a minimum of 70 (35 per group) in pilot studies [44,45]. In our study, 80 participants (n=47, 59% male) were recruited and screened. All (80/80, 100%) were eligible and, thus, enrolled in the study. Informed consent and background information (ie, age and sex) were obtained from each participant. Most of the participants were randomly assigned to either the fixed or adaptive group, whereas efforts were made to balance the sex and age distribution between the 2 groups when we placed the participants in the groups toward the end. The fixed group included 48% (38/80) of the participants (23/38, 61% male), and the adaptive group included 52% (42/80) of the participants (24/42, 57% male). Owing to the convenience of recruitment and proximity to the study location, most participants were students from various departments and programs across the TAMU College Station campus, whereas several faculty and staff members, alumni of the university, and a few residents also participated. As a result, 82% (66/80) of the participants were aged between 18 and 30 years, 9% (7/80) were aged between 31 and 40 years, 2% (2/80) were aged between 41 and 50 years, and 6% (5/80) were aged between 51 and 60 years. There were no participants aged >60 years. Age mean, SD, and IQR; age distribution; and sex distribution are summarized in Table 2. We applied a chi-square test at a 95% confidence level to determine if there were differences in sex and age distribution between the 2 groups. The results showed no difference in the proportions of male, female, and intersex participants in the groups ($\chi^2=0.1, P=.76$) and no difference in the proportion of age in the groups ($\chi^2=2.4, P=.50$).

The administration order between WAIS-IV and e-Cube was randomized. The order of the 3 subtests of the WAIS-IV followed the standardized protocol (BD, DS, and MR), whereas the order of the 6 e-Cube games was randomized. Upon the completion of both tests, the SUS was administered to each participant. The entire session took approximately 90 minutes: 50 minutes for e-Cube, 25 minutes for the WAIS-IV subtests, 5 minutes for the SUS, and a 10-minute break between e-Cube and the WAIS-IV subtests. Each participant was given a US $10 gift card upon the completion of participation.

### Scoring System

The e-Cube games have not been standardized yet, and therefore, scoring methods are not finalized at this stage. We benchmarked the scoring methods used for the WAIS-IV subtests and our previous study [26] and modified them to suit the e-Cube games. The scoring of Assembly considers correctness, item size, and completion time—a 2 × 2 item that is correctly completed within 15 seconds or between 15 and 30 seconds yields 3 or 2 points, respectively; a 3 × 3 item correctly completed within 30 seconds, between 30 and 40 seconds, or between 40 and 60 seconds results in 4, 3, or 2 points, respectively. Shape-Matching, Sequence-Memory, and Spatial-Memory use correctness only as the scoring criteria—2 points for each correct answer and 0 for an incorrect answer. Scoring methods for Path-Tracking and Maze are based on correctness, completion time, and whether the path taken is the shortest. For Path-Tracking, the shortest path finished within 20 seconds, between 20 and 40 seconds, or between 40 and 80 seconds yields 4, 2, or 1 points, respectively; a correct path, but not the shortest, completed within 20 seconds or between 20 and 40 seconds yields 2 or 1 points, respectively. For Maze, the shortest path completed

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants</th>
<th>Fixed group (n=38)</th>
<th>Adaptive group (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; IQR)</td>
<td>26.71 (9.24; 22.00-28.00)</td>
<td>25.74 (8.50; 20.00-27.25)</td>
<td></td>
</tr>
<tr>
<td>Age range (years), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>31 (39)</td>
<td>35 (44)</td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>4 (5)</td>
<td>3 (4)</td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>3 (4)</td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td>&gt;60</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (29)</td>
<td>24 (30)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15 (19)</td>
<td>18 (22)</td>
<td></td>
</tr>
<tr>
<td>Intersex</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Participant demographic data (N=80).
within 10 seconds, between 10 and 20 seconds, or between 20 and 40 seconds yields 4, 2, or 1 points, respectively; a correct path, but not the shortest, completed within 10 seconds or between 10 and 20 seconds yields 2 or 1 points, respectively. Others not satisfying the aforementioned conditions result in 0 points.

The adaptive e-Cube games require some additional considerations for scoring. If an item is generated with the same play complexity as the previous one answered incorrectly, the score for the correct answer is 1 point less than the score used in the fixed version. A total of 2 consecutive incorrect answers result in the system generating an easier item, and in this case, a correct answer for that newly generated item yields only 1 point.

Statistical Analysis
Correlations were computed to determine the relationships between the computed complexity values and participants’ performance, the connections between the WAIS subtests and the e-Cube games, and the relationships among the 6 e-Cube games. We used the Spearman correlation to measure the monotonic association among them. The correlation is interpreted as “weak,” “moderate,” and “strong/high” when the coefficient is <0.36, between 0.36 and 0.67, and >0.67, respectively [46]. We used 2-tailed $t$ tests to identify the mean differences in the game or subtest scores and the SUS scores between the 2 groups.

Ethics Approval and Informed Consent
This human participant study was reviewed and approved by the TAMU Institutional Review Board (IRB2019-1079D; approval date: December 22, 2020). Informed consent was obtained from all participants before taking part in this study.

Results
All enrolled participants (80/80, 100%) completed the entire session without withdrawal. The results and findings for objectives 1 and 2 are presented in the following sections.
Table 3. Correlations (Spearman $r$, 2-tailed $P$ value, and 95% CIs) between the $C_{play}$ values and the mean correctness and mean completion time for each item from the fixed group participants.

<table>
<thead>
<tr>
<th>Game (df)</th>
<th>Mean completion time</th>
<th>Mean correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.86 $^a$</td>
<td>−0.50</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.02</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.67 to 0.94</td>
<td>−0.46 to 0.43</td>
</tr>
<tr>
<td>Shape-Matching (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.95</td>
<td>−0.75</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.009</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.80 to 0.99</td>
<td>−0.94 to −0.23</td>
</tr>
<tr>
<td>Sequence-Memory (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.98</td>
<td>−0.95</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.94 to 0.99</td>
<td>−0.98 to −0.86</td>
</tr>
<tr>
<td>Spatial-Memory (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.94</td>
<td>−0.30</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.30</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.76 to 0.99</td>
<td>−0.78 to 0.41</td>
</tr>
<tr>
<td>Path-Tracking (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.82</td>
<td>−0.20</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.002</td>
<td>.58</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.39 to 0.96</td>
<td>−0.74 to 0.49</td>
</tr>
<tr>
<td>Maze (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.72</td>
<td>−0.59</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.01</td>
<td>.07</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.17 to 0.93</td>
<td>−0.89 to 0.06</td>
</tr>
</tbody>
</table>

$^a$Italics indicate that a correlation existed.

Objective 2: Evaluation of Preliminary Utility and Usability of e-Cube Games for Cognitive Assessment

Overview

The mean, SDs, and IQR values obtained from participants in each group for the WAIS-IV subtests (raw scores) and e-Cube games are summarized in Table 4. We also conducted a 2-tailed $t$ test with equal variance (Cronbach $\alpha$=.05) comparing the test scores from the adaptive and fixed groups to determine whether significant differences existed in mean scores between the 2 groups. The $t$ test showed no significant differences in the mean scores of the 3 WAIS subtests—BD ($P$=.37), MR ($P$=.06), and DS ($P$=.18)—between the 2 groups. Group differences were found in Shape-Matching and Sequence-Memory, but not in other e-Cube games.
Table 4. Score statistics from the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), subtests and e-Cube games.

<table>
<thead>
<tr>
<th></th>
<th>Fixed group, mean (SD; IQR)</th>
<th>Adaptive group, mean (SD; IQR)</th>
<th>2-tailed t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WAIS-IV raw score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BD&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50.95 (11.26; 41.00-60.00)</td>
<td>53.02 (9.35; 47.75-60.00)</td>
<td>-0.90 (78)</td>
<td>.37</td>
</tr>
<tr>
<td>DS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>28.34 (4.86; 24.75-32.00)</td>
<td>29.98 (5.83; 26.00-34.00)</td>
<td>-1.36 (78)</td>
<td>.18</td>
</tr>
<tr>
<td>MR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>21.55 (2.45; 11.00-14.00)</td>
<td>22.62 (2.49; 21.00-24.00)</td>
<td>-1.93 (78)</td>
<td>.06</td>
</tr>
<tr>
<td><strong>e-Cube score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assembly</td>
<td>54.37 (11.10; 48.75-62.00)</td>
<td>58.52 (12.52; 51.75-67.25)</td>
<td>-1.56 (78)</td>
<td>.12</td>
</tr>
<tr>
<td>Shape-Matching</td>
<td>16.53 (1.84; 16.00-18.00)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>14.67 (2.86; 13.00-16.25)</td>
<td>3.42 (78)</td>
<td>.001</td>
</tr>
<tr>
<td>Sequence-Memory</td>
<td>20.63 (4.63; 16.00-24.00)</td>
<td>17.98 (3.83; 15.00-20.25)</td>
<td>2.81 (78)</td>
<td>.006</td>
</tr>
<tr>
<td>Spatial-Memory</td>
<td>17.21 (2.16; 16.00-18.50)</td>
<td>16.88 (2.93; 15.00-19.00)</td>
<td>0.57 (78)</td>
<td>.57</td>
</tr>
<tr>
<td>Path-Tracking</td>
<td>26.95 (5.83; 24.00-31.00)</td>
<td>26.21 (7.54; 22.00-31.25)</td>
<td>0.48 (78)</td>
<td>.63</td>
</tr>
<tr>
<td>Maze</td>
<td>22.47 (6.27; 18.75-26.25)</td>
<td>22.12 (5.18; 17.00-26.00)</td>
<td>0.28 (78)</td>
<td>.78</td>
</tr>
</tbody>
</table>

<sup>a</sup>BD: Block Design.
<sup>b</sup>DS: Digit Span.
<sup>c</sup>MR: Matrix Reasoning.
<sup>d</sup>Italics indicate that a difference existed.

Relationship Between e-Cube Games and WAIS-IV Subtests

We presented the expected relationships between the e-Cube games and the WAIS-IV subtests in Table 1. The evaluation results are shown in Tables 5 and 6, which list the correlations between the e-Cube scores and WAIS-IV subtest scores in the fixed and adaptive groups, respectively. The 2 groups showed somewhat different trends in results. In both groups, Assembly and BD were moderately correlated, as expected in Table 1. Shape-Matching was expected to be correlated with MR, and the results from the adaptive group agreed with this. The Shape-Matching results from the fixed group showed a weak correlation with BD but no correlation with MR. Sequence-Memory and Spatial-Memory were expected to tap working memory as assessed by DS, but only the adaptive version of Spatial-Memory was moderately correlated with DS. The adaptive version of Sequence-Memory and BD also showed a weak correlation. Path-Tracking and Maze were expected to be related to BD and MR, and only the adaptive version of Path-Tracking yielded the expected results. However, no correlations were found between both versions of Maze and any WAIS subtest. Another notable finding was that both versions of Sequence-Memory showed no significant correlation with DS. Overall, the results suggested that the adaptive versions better tap into the cognitive abilities assessed by the 3 WAIS-IV subtests.

We further analyzed the intercorrelations among the e-Cube games (Multimedia Appendix 1 for fixed games and Multimedia Appendix 2 for adaptive games). The adaptive e-Cube showed fewer intercorrelations than the fixed version. In the fixed version, most of the games were somewhat correlated except for Shape-Matching. Both versions of Path-Tracking and Maze were correlated with Assembly.
Table 5. Correlations (Spearman $r$, 2-tailed $P$ value, and 95% CIs) between the e-Cube scores and raw scores of the Wechsler Adult Intelligence Scale, Fourth Edition, subtests in the fixed group.

<table>
<thead>
<tr>
<th>Fixed game</th>
<th>BD$^a$</th>
<th>DS$^b$</th>
<th>MR$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.49^{d,e}$</td>
<td>$-0.20$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.002^{d}$</td>
<td>$0.23$</td>
<td>$0.57$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$0.20$ to $0.70^{d}$</td>
<td>$-0.49$ to $0.13$</td>
<td>$-0.23$ to $0.41$</td>
</tr>
<tr>
<td>Shape-Matching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.33$</td>
<td>$-0.30$</td>
<td>$0.09^{d}$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.04$</td>
<td>$0.07$</td>
<td>$0.57^{d}$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$0.01$ to $0.59^{d}$</td>
<td>$-0.57$ to $0.02$</td>
<td>$-0.24$ to $0.40^{d}$</td>
</tr>
<tr>
<td>Sequence-Memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.31$</td>
<td>$0.14^{d}$</td>
<td>$0.26$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.06$</td>
<td>$0.40^{d}$</td>
<td>$0.11$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$-0.01$ to $0.57$</td>
<td>$-0.19$ to $0.44^{d}$</td>
<td>$-0.07$ to $0.54$</td>
</tr>
<tr>
<td>Spatial-Memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.26$</td>
<td>$0.23^{d}$</td>
<td>$0.19$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.12$</td>
<td>$0.16^{d}$</td>
<td>$0.25$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$-0.07$ to $0.54$</td>
<td>$-0.10$ to $0.51^{d}$</td>
<td>$-0.14$ to $0.48$</td>
</tr>
<tr>
<td>Path-Tracking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.43^{d}$</td>
<td>$0.04$</td>
<td>$0.04^{d}$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.007^{d}$</td>
<td>$0.81$</td>
<td>$0.80^{d}$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$0.13$ to $0.66^{d}$</td>
<td>$-0.28$ to $0.36$</td>
<td>$-0.28$ to $0.36^{d}$</td>
</tr>
<tr>
<td>Maze</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.30^{d}$</td>
<td>$-0.05$</td>
<td>$0.17^{d}$</td>
</tr>
<tr>
<td>$P$ value</td>
<td>$0.06^{d}$</td>
<td>$0.78$</td>
<td>$0.29^{d}$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$-0.02$ to $0.57^{d}$</td>
<td>$-0.36$ to $0.27$</td>
<td>$-0.16$ to $0.46^{d}$</td>
</tr>
</tbody>
</table>

$^a$BD: Block Design.  
$^b$DS: Digit Span.  
$^c$MR: Matrix Reasoning.  
$^d$Indicates that the 2 were expected to be correlated in Table 1.  
$^e$Italics indicate that a correlation existed.
Table 6. Correlations (Spearman r, 2-tailed P value, and 95% CI) between the e-Cube scores and raw scores of the Wechsler Adult Intelligence Scale, Fourth Edition, subtests in the adaptive group.

<table>
<thead>
<tr>
<th>Adaptive game</th>
<th>BD&lt;sup&gt;a&lt;/sup&gt;</th>
<th>DS&lt;sup&gt;b&lt;/sup&gt;</th>
<th>MR&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.49&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>0.18</td>
<td>0.31</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.25</td>
<td>.05</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.21 to 0.70&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.14 to 0.47</td>
<td>−0.00 to 0.57</td>
</tr>
<tr>
<td>Shape-Matching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.26</td>
<td>0.22</td>
<td>0.34&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.09</td>
<td>.16</td>
<td>.03&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>95% CI</td>
<td>−0.06 to 0.53&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.10 to 0.50</td>
<td>0.03 to 0.59&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sequence-Memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.34</td>
<td>0.21&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.11</td>
</tr>
<tr>
<td>P value</td>
<td>.03</td>
<td>.19&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.50</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.03 to 0.59&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.11 to 0.49&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.21 to 0.41</td>
</tr>
<tr>
<td>Spatial-Memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.09</td>
<td>0.51&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.17</td>
</tr>
<tr>
<td>P value</td>
<td>.59</td>
<td>&lt;.001&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.29</td>
</tr>
<tr>
<td>95% CI</td>
<td>−0.23 to 0.39&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.24 to 0.71&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.15 to 0.46</td>
</tr>
<tr>
<td>Path-Tracking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.45&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.01</td>
<td>0.45&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.003&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.93</td>
<td>.003&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.16 to 0.67&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.32 to 0.30</td>
<td>0.16 to 0.67&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Maze</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.30&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.07</td>
<td>0.11&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.05&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.62</td>
<td>.49&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>95% CI</td>
<td>−0.01 to 0.56&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−0.25 to 0.37</td>
<td>−0.21 to 0.41&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>BD: Block Design.
<sup>b</sup>DS: Digit Span.
<sup>c</sup>MR: Matrix Reasoning.
<sup>d</sup>Indicates that the 2 were expected to be correlated in Table 1.
<sup>e</sup>Italics indicate that a correlation existed.

### Technical Reliability and Usability of e-Cube

The e-Cube technology operated smoothly without any substantial technical issues identified during the study. The false detection rate, defined as the percentage ratio of incorrect detections to the total number of detections, was approximately 0.1% (6/5990). We note that all the analyses and computations mentioned previously were based on the corrected data. To analyze the results of the SUS, the scores from the 10 items were converted into a scale of 0 to 100 [47]. The overall mean SUS score was 83.40 (SD 11.52). There was a significant group difference in the results. The mean of the SUS scores from the fixed group participants was 80.79 (SD 13.23), whereas the mean score from the adaptive group participants was 86.01 (SD 8.75). The 2-sample, 2-tailed t test with Cronbach α=.05 showed t<sub>78</sub>=−2.10 (P=.04). The result from the adaptive group showed a significantly higher mean SUS score with a smaller SD than the fixed group. On the basis of the industry standard [48], the usability of both fixed and adaptive e-Cube games is considered grade A (ie, the games are acceptable).
Discussion

Principal Findings

We presented the design, development, and evaluation of the e-Cube system for automated cognitive assessment. e-Cube is a vision-based system converted from TAG-Games, a computerized system using a set of highly instrumented blocks [28]. e-Cube adopted a set of plastic cubes and a webcam instead, costing only approximately US $50. e-Cube also reduced the labor burden by generating adaptive items, detecting answers and behavior, and scoring autonomously. A total of 6 games—Assembly, Shape-Matching, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze—were designed using the proposed play complexity measures and adaptive item generators. The e-Cube technology and the adaptive games were evaluated by testing the 2 objectives. This human participant study was conducted on the TAMU campus, and thus, most of our study participants (66/80, 82%) were college students aged between 18 and 30 years, with only 18% (14/80) aged between 31 and 60 years. Therefore, the results must be interpreted considering the skewed age distribution and demographic characteristics.

Objective 1 was supported by the correlation analyses performed between the $C_{play}$ values and the 2 performance indicators—the mean correctness and mean completion time obtained from the fixed group. We found that each game was correlated with at least one performance indicator. The $C_{play}$ values of Assembly, Shape-Matching, and Sequence-Memory showed high correlations with both means. No correlations were found using the mean correctness in Spatial-Memory, Path-Tracking, and Maze. As discussed previously, correctness was not the dominant factor that widened the performance difference in Path-Tracking and Maze; thus, no correlations with correctness were found. For Spatial-Memory, 2 items involved symmetric arrangements of the geometric shapes—which made them easy to memorize regardless of the geometric complexity of the shapes. We used the same play complexity measure for both Spatial-Memory and Assembly, which appeared not to ideally reflect the difficulty associated with such memory tasks despite a high correlation in the mean completion time. This problem can be avoided at the software level by adjusting the algorithm for the item generator. Nevertheless, such symmetric images were rarely created in the adaptive version and, thus, are expected to have minimal effect on the assessment outcome.

To test the preliminary utility of the adaptive e-Cube games for cognitive assessment (objective 2), correlation analyses were performed between the scores from the e-Cube games and the WAIS-IV subtests. The adaptive games yielded more significant correlations with the WAIS-IV subtests than the fixed ones. This implies the potential utility of the adaptive feature of the e-Cube games based on the complexity measures. The adaptive version used a discontinuation rule (ie, the substantially small step size leading to the termination of the game), which possibly reduced the number of items in each game, fatigue, and unintended correct answers. For example, given a fixed number of items sorted by increasing difficulties, one may fail to answer correctly in the early items but can unintentionally provide correct answers in the later items. The automatic item generator in the adaptive games adjusts the item complexity based on real-time performance, enabling the system to generate a more appropriate assessment for everyone. Note that the administration of the WAIS-IV also applies the discontinuation rule in the subtests to minimize time [49]. The subtest is terminated when a participant fails to answer a certain number of consecutive items, which differs for each subtest. Intercorrelation analyses also showed that the games in the adaptive version were more independent of one another.

There were some other notable findings from the objective 2 evaluation study. The mean scores of Shape-Matching and Sequence-Memory in the fixed group were higher than those in the adaptive group (Table 4). In the fixed group, we found that most of the participants (26/38, 68%) correctly answered items 1 to 7 and 9, whereas only 50% (19/38) answered item 8 correctly and 24% (9/38) answered item 10 correctly in Shape-Matching. This inconsistency resulted in a higher mean score in the fixed group, but the results were not correlated with MR. In contrast, the adaptive group showed a significant correlation between Shape-Matching and MR. Sequence-Memory and Spatial-Memory were evaluated to understand which game has a monotonic relationship with DS, but a correlation was only found between the adaptive version of Spatial-Memory and DS. DS measures verbal working memory, which relies on auditory recall of numbers, sequences, and orders. However, Sequence-Memory was performed through the visual recall of geometric images, and Spatial-Memory used visual-spatial images. This fundamental design difference may have led to a lack of correlation. In addition, the language differences in participants and how they differently affect DS scores were not analyzed in this study as we did not collect such background data. Some nonnative speakers mentioned slight difficulty in memorizing the numbers said in English during the DS subtest. This feedback was collected only informally. The Path-Tracking and Maze games were correlated with Assembly, implying that their game settings or measured cognitive outcomes were similar to those of Assembly. Furthermore, Maze was not correlated with any WAIS subtest.

We further analyzed the correlation between the composite scores of the 6 e-Cube games and those of the 3 WAIS subtests. The results were $r_{38}=0.51$ (95% CI 0.23-0.71; $P<.001$) for the fixed group and $r_{38}=0.53$ (95% CI 0.26-0.72; $P<.001$) for the adaptive group. When only 4 e-Cube games (ie, Assembly, Shape-Matching, Sequence-Memory, and Spatial-Memory) were considered, the results were $r_{38}=0.50$ (95% CI 0.21-0.71; $P<.001$) for the fixed group and $r_{38}=0.59$ (95% CI 0.34-0.76; $P<.001$) for the adaptive group. Path-Tracking and Maze did not result in any meaningful relationship with the WAIS, and thus, their potential utility in cognitive assessment requires further exploration.

The low false detection rate (0.1%) demonstrated the technical functionality of the e-Cube system. Regarding the usability evaluation (objective 2), the average SUS scores from participants in both the fixed and adaptive groups were acceptable based on the industry standard [48]. The adaptive games resulted in a considerably higher mean SUS score with...
a smaller SD than that of the fixed games. To understand the feedback for individual items, we combined the results from both groups and computed the average rate for each item. The results for the individual SUS items were uniformly positive. For the 5 even-numbered questions that were in a negative tone, such as “I found the e-Cube games unnecessarily complex,” all rates were between 1 (strongly disagree) and 2 (disagree). For the odd-numbered questions that were in a positive tone, the rates were between 4 (agree) and 5 (strongly agree) except for the following question—“I think that I would like to use the e-Cube games frequently”—with an average rate of 3.8. This was mainly due to the e-Cube games taking relatively long (approximately 50 minutes) to complete at this preliminary stage. Most of our participants (73/80, 91%) were aged <40 years, so using the game frequently to track cognitive decline was unnecessary for them. We received the highest evaluation of 4.5 on the following item: “I found the various functions in the e-Cube games were well integrated among all questions.”

Limitations

Most participants (66/80, 82%) were TAMU students aged between 18 and 30 years. The data from participants who were young, educated, and motivated do not represent the general population well. This may also explain why none of the participants withdrew from the study. Furthermore, additional demographic information such as education level, socioeconomic status, race, and ethnicity was not collected in this preliminary evaluation study. A larger-scale validation study will be needed to involve participants from various communities with diverse backgrounds.

The administration order of the 6 e-Cube games was randomized to control for an order effect. The test order can influence the test results and bring about different levels of fatigue [50], so a well-developed cognitive assessment usually requires a standardized administration order. Although order and fatigue effects were not found in some standardized tests [50,51], the impact of the administration order of e-Cube on the scores was not investigated.

The WAIS-IV DS includes Forward, Backward, and Sequencing, which measure auditory working memory and attention with information reordering. However, Sequence-Memory and Spatial-Memory rely on visual recall and do not require any manipulation of information. Therefore, DS may not be an ideal choice for validating these 2 games. Another measure, such as Spatial Span Forward in the Wechsler Memory Scale, Fourth Edition, may be selected to compare the results with those of Sequence-Memory and Spatial-Memory in measuring relevant working memory skills.

Future Work

The e-Cube technology was developed for fully autonomous administration and scoring of cognitive assessment targeting fine motor, hand-eye coordination, cognitive reasoning, and working memory skills. Building on our prior work [26,28], the technology was converted into a much simpler, cheaper, and easy-to-use form, thus showing potential for use in larger-scale research studies. Our long-term objective is for the e-Cube games to serve as a routine self-assessment tool used by individuals who require continuous monitoring of their cognitive health, such as older adults and people with mild cognitive impairment. Once fully established, e-Cube can also be adapted in clinical settings, especially for remote assessment without requiring in-person interactions with an administrator. Our future work will be geared toward this long-term objective.

This extended human participant study will involve diverse participants (eg, age, sex, education, and socioeconomic status) to better represent the general population. In this future evaluation study, the existing instruments for comparison must be revisited and selected to ensure that the measures match the target cognitive domains of individual e-Cube games. Future work will also aim to establish reliability via test-retest evaluation and the validity of self- and remote administration functions via comprehensive and comparative evaluations. The study to understand the user experiences may also be extended by including an additional set of questionnaires to compare traditional instruments and the e-Cube games to gauge their preference if the e-Cube system is proven to replace some of these. To further improve the technical performance, additional vision processing methods may be added to improve this rate, such as hand detection algorithms to prevent hand motions from interfering with block detection.

The rich data from the e-Cube games on patterns, speed, and characteristics of physical movements applied to the cubes can also be explored to further explicate individual differences and cognitive and fine motor deficits. Such behavioral data may hold important information about individuals, especially those with cognitive deficits exacerbated by fine motor deficits or other behavioral symptoms such as hand tremors. Furthermore, the data provided by e-Cube have the potential to assess one’s cognitive skills in a more objective way than in standard clinical settings. Upon validating its utility as a cognitive assessment tool, our future research may explore the e-Cube games for screening of early signs of neurological diseases. For the e-Cube games to be used as a routine assessment tool, we will consider shortening and gamifying the assessment to make it more fun and engaging. Enhanced graphics and sound and visual feedback mechanisms may be added to the game design. For example, we may benchmark the features of Music Blocks and iSIG-Blocks [52,53], allowing the users to customize audio, tactile, and visual sensory feedback during the cognitive assessment. The current system runs on a low-end laptop with a webcam, whereas further developments can make the algorithms executable on a tablet or cell phone using their built-in cameras. This may further reduce the cost, make it suitable for self- or remote assessment, and support long-term adoption and broader use of the technology.
Acknowledgments
This work was supported by the National Science Foundation under grant 2002721. The authors thank Elisabeth S Ford for assisting with the human participant study and all study participants for their time and contribution to this project.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Intercorrelations (Spearman r_{38}, 2-tailed P value, and 95% CIs) within the scores of fixed games.

References


27. Jeong DH. Distributed wireless sensor network systems: theoretical framework, algorithms, and applications. Case Western Reserve University. 2015. URL: https://etd.ohiolink.edu/apexprod/rws_olk/1/1501/10?clear=10&pl0_accession_num=case1436541959 [accessed 2023-04-17]


Abbreviations

BD: Block Design
DS: Digit Span
GLCM: gray-level co-occurrence matrix
MR: Matrix Reasoning
SUS: System Usability Scale
TAMU: Texas A&M University
WAIS: Wechsler Adult Intelligence Scale
WAIS-IV: Wechsler Adult Intelligence Scale, Fourth Edition
Advantages of a Training Course for Surgical Planning in Virtual Reality for Oral and Maxillofacial Surgery: Crossover Study

Max Ulbrich¹; Vincent Van den Bosch², MD; Andrea Bönsch³, MSc; Lennart Johannes Gruber¹, DMD; Mark Ooms¹, MD; Claire Melchior¹; Ila Motmaen¹, DMD; Caroline Wilpert², MD; Ashkan Rashad¹, MD, DMD; Torsten Wolfgang Kuhlen³, PhD; Frank Hölzle¹, PhD, MD; Behrus Puladi¹,², MD, DMD

¹Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Aachen, Germany
²Department of Diagnostic and Interventional Radiology, University Hospital RWTH Aachen, Aachen, Germany
³Visual Computing Institute, Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen University, Aachen, Germany
⁴Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany

Corresponding Author:
Behrus Puladi, MD, DMD
Department of Oral and Maxillofacial Surgery
University Hospital RWTH Aachen
Pauwelstraße 30
Aachen, 52074
Germany
Phone: 49 241 80 88231
Fax: 49 241 80 82430
Email: bpuladi@ukaachen.de

Abstract

Background: As an integral part of computer-assisted surgery, virtual surgical planning (VSP) leads to significantly better surgery results, such as for oral and maxillofacial reconstruction with microvascular grafts of the fibula or iliac crest. It is performed on a 2D computer desktop screen (DS) based on preoperative medical imaging. However, in this environment, VSP is associated with shortcomings, such as a time-consuming planning process and the requirement of a learning process. Therefore, a virtual reality (VR)-based VSP application has great potential to reduce or even overcome these shortcomings due to the benefits of visuospatial vision, bimanual interaction, and full immersion. However, the efficacy of such a VR environment has not yet been investigated.

Objective: This study aimed to demonstrate the possible advantages of a VR environment through a substep of VSP, specifically the segmentation of the fibula (calf bone) and os coxae (hip bone), by conducting a training course in both DS and VR environments and comparing the results.

Methods: During the training course, 6 novices were taught how to use a software application in a DS environment (3D Slicer) and in a VR environment (Elucis) for the segmentation of the fibula and os coxae, and they were asked to carry out the maneuvers as accurately and quickly as possible. Overall, 13 fibula and 13 os coxae were segmented for each participant in both methods (VR and DS), resulting in 156 different models (78 fibula and 78 os coxae) per method (VR and DS) and 312 models in total. The individual learning processes in both environments were compared using objective criteria (time and segmentation performance) and self-reported questionnaires. The models resulting from the segmentation were compared mathematically (Hausdorff distance and Dice coefficient) and evaluated by 2 experienced radiologists in a blinded manner.

Results: A much faster learning curve was observed for the VR environment than the DS environment ($\beta=.86$ vs $\beta=.25$). This nearly doubled the segmentation speed ($\text{cm}^3/\text{min}$) by the end of training, leading to a shorter time ($P<.001$) to reach a qualitative result. However, there was no qualitative difference between the models for VR and DS ($P=.99$). The VR environment was perceived by participants as more intuitive and less exhausting, and was favored over the DS environment.

Conclusions: The more rapid learning process and the ability to work faster in the VR environment could save time and reduce the VSP workload, providing certain advantages over the DS environment.

(JMIR Serious Games 2023;11:e40541) doi:10.2196/40541
Introduction

Malignant or benign tumors, advanced osteomyelitis, osteoradionecrosis, and complex fractures can lead to extensive bone and soft tissue defects with the need for reconstruction. Therefore, soft and hard tissue reconstruction is an important and commonly used procedure in oral and maxillofacial surgery [1]. Microvascular reconstruction with fibula or iliac crest bone transplants is one of the best options for mandible bone reconstruction [2,3], and it has the highest success rate and delivers the best functional and esthetic results [4,5]. Microvascular reconstruction has been performed using a conventional technique, specifically manual free transplant raising, adjustment, and insertion [6,7].

Given its numerous clinical advantages, a virtual surgical planning (VSP) approach prior to microvascular reconstruction of the jaw has seen increased popularity. This approach involves using preoperative medical imaging within computer-assisted surgery (CAS) applications [8,9]. The advantages of VSP include reduced ischemia, a shorter defect reconstruction time, a shorter surgical procedure [10], a shorter length of hospital stay [10], a lower number of necessary osteotomy revisions, an overall lower volume of bone removed, a lower rate of osseous injury [11], and a better match of removed bone volume to defect volume [12].

However, VSP is associated with a higher preoperative workload [8] and is therefore often delegated to younger fellows or technical staff [13,14]. Despite this, VSP still remains an integral part of the surgical process and should be done or supervised by the performing surgeon [15]. Furthermore, learning VSP is time-consuming and requires an appropriate amount of learning time, but these investments are required to ultimately achieve good clinical outcomes [15-18].

One of the main bottlenecks is the preparation of 3D models by segmentation [19]. This process is still performed on a 2D computer desktop screen (DS) with 2D controls, such as a mouse and keyboard [20], which however seems unsuitable per se for such a task [21]. Furthermore, working in a DS environment differs from working in a surgical site since the DS environment lacks stereoscopic vision, and the use of a mouse and keyboard does not resemble working with surgical instruments at all. This leads to discrepancies between the VSP performed in a DS environment and the surgery performed in an operating theater.

A potential way to overcome the disadvantages of VSP performed in a DS environment could be the use of virtual reality (VR) [22] since it offers stereoscopic vision, allows the user to work manually, and allows more focused work due to immersion [21,23]. VR environments better resemble realistic work routines in the operating theater, as users can rotate and flip anatomical structures and observe different intraoperative viewing angles [24], thus enabling an enhanced understanding of the anatomy [25]. Although VR more broadly has been applied to surgery, it has often been basic or partially immersive VR. This must be distinguished from enhanced, fully immersive, binocular head-based VR [26]. With the so-called second wave that began in 2012, this technology has significantly spread in the consumer market and must be distinguished from augmented reality–based systems, which have become widespread under the terms mixed or extended reality [27]. Areas of application for these new VR head-mounted displays (HMDs) are surgical education, surgical training, and surgical planning [28]. VR simulations have been shown to lead to an improvement in surgical skills among subjects [29]. VR was also used for the visualization of medical images in radiology [22]. Recently, VR has been used for spatial understanding [30] or in the course of multiuser conferences during surgical planning [31]. However, to our knowledge, the potential for performing all steps or substeps of VSP using VR HMDs has not yet been explored.

This study aimed to demonstrate the possible advantages of a VR environment through a substep of VSP, specifically the segmentation of the fibula (calf bone) and os coxae (hip bone), which is typically applied in oral and maxillofacial surgery, by conducting a training course in both DS and VR environments and comparing the results.

Methods

Cases

We retrospectively selected 78 (13 cases for every 6 participants) planning computed tomography (CT) scans acquired between 2015 and 2020 at the University Hospital RWTH Aachen originally intended for VSP for the mandible, maxilla, or other viscerocranial sites for microvascular reconstruction. The CT scans were first scored by 2 trained radiologists based on image quality (1, good; 2, moderate; 3, poor), bone quality (1, good; 2, moderate; 3, poor), and artifacts (1, none; 2, moderate; 3, plenty), and from this, a total score (range 3-9) was then derived for each case and radiologist. A mean value was then calculated for each case based on the 2 radiologists’ total scores.

Ethics Approval

The local ethics committee of the Medical Faculty of RWTH Aachen University approved our study (approval number EK 471/20). The experimental protocol was carried out in accordance with the guidelines set by the Declaration of Helsinki. Informed consent was obtained from all participants involved in the study.

Study Design

A total of 6 novices in VSP (5 oral and maxillofacial surgery residents and 1 final-year dentistry student) took part in the study. Each participant’s age (mean 33.2, SD 2.6 years), gender (1 female, 5 males), surgical work experience (0-4 years), prior experience with computers, and prior experience with VR were recorded. Each participant performed a mental rotation test [32] before training to examine the influence of their baseline visuospatial ability on segmentation performance.
All novices received an introduction to the basic principles of the VSP application. The participants were stratified into the following 2 groups based on surgical work experience: Group A (4 years, 3 years, and 0 years in residency) and Group B (4 years, 2 years, and 1 year in residency). Group A was trained with 13 (3 cases for initial warm-up and 10 cases for autonomous training) randomly selected (for a realistic clinical scenario) cases per participant, first in the DS environment and thereafter in the VR environment. In contrast, Group B was trained with 13 randomly selected cases per participant, first in the VR environment and thereafter in the DS environment (Figure 1). This ensured that a potential learning effect gained in one environment and transferred to the other was evenly distributed between both [18,33]. The given task of the training was the segmentation of the fibula (a simpler model) and os coxae (a more complex model) of the right side of the body in the corresponding working environment. The participants were asked to carry out their tasks as accurately and quickly as possible. In the case of relevant bone defects, the left side was used.

**Figure 1.** Graphic illustration of the study protocol. A total of 78 computed tomography scans were included, scored by 2 radiologists for image quality, and randomized to 6 participants, who additionally performed a mental rotation test before training. Subsequently, the 6 participants were stratified into Groups A and B, each of which started with the virtual reality (VR) or desktop screen (DS) environment, respectively. During the training, Likert-type questionnaires were filled out, and segmentation time was measured. After training, the System Usability Scale (SUS), the User Experience Questionnaire (UEQ), and a final Likert-type questionnaire were administered. After training, all fibula and hip bone models were compared using the Hausdorff distance and the Dice coefficient, and additionally evaluated in a blinded fashion by the same 2 radiologists.

Each participant received an appropriate briefing in the corresponding working environment (DS or VR) with an explanation of all important functions, whereby the standard analog functions of both applications had to be used within a...
test case. Subsequently, the task was carried out based on a test case under the supervision of an experienced user who offered verbal guidance throughout each work step. The participants then started to work on the 13 cases. The opportunity to ask questions was allowed only for the first 3 cases to ensure a realistic clinical scenario. After each case, the procedure time was recorded (for fibula and os coxae segmentation, respectively), and a Likert-type questionnaire was completed [34]. All participants successfully completed the training. After completing the training for both environments, a final Likert-type questionnaire was completed. Additionally, for each application, the User Experience Questionnaire (UEQ) [35] and System Usability Scale (SUS) [36] were administered.

DS-based training was performed using 3D Slicer version 4.11.20210226 [37], and VR-based training was performed using Elucis version 1.4 (Realize Medical Inc). The VR hardware included an HTC Vive Pro with an HTC Vive Controller 2.0 (Valve Corporation). All training was done seated (Figures 2-5) at the same workstation (AMD Ryzen 3900X CPU with 64 GB of memory and an RTX 2080 Ti graphics card).

**Figure 2.** The virtual reality working environment (Elucis) and a segmented hip bone model in yellow in the middle.

**Figure 3.** The virtual reality working environment from a third-person perspective.
Evaluation
All segmented models of the fibula and os coxae (n=156) were assessed in a blinded setting by the same 2 trained radiologists using an absolute category rating (1, excellent; 2, good; 3, fair; 4, poor; 5, bad). Based on the evaluation of both radiologists, a mean value was calculated.

Afterwards, the postprocessing of all cases was performed using the 3D Slicer add-on Surface Wrap Solidify [38] to remove cavities in the models. This was necessary to avoid bias when comparing the volumes or surfaces of the models. Hausdorff distances and Dice coefficients [39] were then computed using the 3D Slicer add-on segment comparison [40] between the DS model and the VR model. Finally, intersection volumes (cm$^3$) between the VR and DS models were divided by segmentation duration (minutes) to calculate common segmentation performance (cm$^3$/min).

Statistical Analysis
Statistical analysis was performed using the programming language R (Version 4.1.1; R Foundation for Statistical Computing). A $P$ value <.05 was considered significant. We used a $t$ test, a Wilcoxon signed-rank test, or a chi-square test to assess differences between VR and DS. The training effect as a function of the environment was determined as the duration of task completion using a linear mixed effect model with the
R package lme4. For the model-based evaluations, a likelihood ratio test of the corresponding parameters was used to evaluate the relationships between dependent and independent variables. *P* values were adjusted for multiple testing using the Holm-Bonferroni method [41]. The 95% CIs were calculated by conventional bootstrapping with 1000 replications [42]. Plots were visualized using the R package ggplot2.

**Results**

**Segmentation**

The average overall radiologist ratings for the segmentation results were 1.40 (1, excellent; 5, bad) for the VR models and 1.29 for the DS models, and were not significantly different (Wilcoxon signed-rank test, *P*=.29). Similarly, there was no significant difference in radiologist assessment for the os coxae (VR 1.62 vs DS 1.39; Wilcoxon signed-rank test, *P*=.14) or fibula (VR 1.17 vs DS 1.17; Wilcoxon signed-rank test, *P*=.99; Table 1). Furthermore, regarding clinical use (models with a score ≤2/good), there was no relevant difference between VR and DS for all models, os coxae models, and fibula models (χ², *P*=.99, *P*=.99, and *P*=.99, respectively). Thus, the VR models are equally suitable for clinical use as the DS models.

Table 1. Results of metric evaluations and blinded assessments by 2 radiologists.

<table>
<thead>
<tr>
<th>Model and characteristics (n=312)</th>
<th>VR&lt;sup&gt;a&lt;/sup&gt; environment, mean (SD)</th>
<th>DS&lt;sup&gt;b&lt;/sup&gt; environment, mean (SD)</th>
<th>Difference, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Os coxae (hip bone)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone volume (cm³)</td>
<td>341 (65)</td>
<td>329 (62)</td>
<td>12 (12)</td>
</tr>
<tr>
<td>Segmentation duration (min)</td>
<td>22.5 (13.0)</td>
<td>38.7 (22.6)</td>
<td>−16.2 (21.3)</td>
</tr>
<tr>
<td>Segmentation performance (cm³/min)</td>
<td>19.2 (12.8)</td>
<td>12.0 (8.6)</td>
<td>7.2 (4.2)</td>
</tr>
<tr>
<td>Segmentation quality (range 1-5)</td>
<td>1.62 (0.76)</td>
<td>1.39 (0.63)</td>
<td>0.23 (0.78)</td>
</tr>
<tr>
<td>Hausdorff distance (mm)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
<td>0.43 (0.19)</td>
</tr>
<tr>
<td>Dice coefficient (%)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.96 (0.02)</td>
</tr>
<tr>
<td><strong>Fibula (calf bone)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone volume (cm³)</td>
<td>60 (12)</td>
<td>58 (11)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Segmentation duration (min)</td>
<td>12.1 (8.0)</td>
<td>17.0 (11.1)</td>
<td>−4.9 (8.3)</td>
</tr>
<tr>
<td>Segmentation performance (cm³/min)</td>
<td>6.5 (2.9)</td>
<td>4.7 (4.0)</td>
<td>1.8 (1.1)</td>
</tr>
<tr>
<td>Segmentation quality (range 1-5)</td>
<td>1.17 (0.38)</td>
<td>1.20 (0.44)</td>
<td>−0.03 (0.52)</td>
</tr>
<tr>
<td>Hausdorff distance (mm)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.29 (0.12)</td>
</tr>
<tr>
<td>Dice coefficient (%)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.96 (0.02)</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.

<sup>b</sup>DS: desktop screen.

<sup>c</sup>N/A: not applicable.

In contrast, the segmentation results from the VR environment were considered better than those from the DS environment by the participants themselves for both the os coxae models (7-point Likert scale [1, strongly disagree; 7, strongly agree], VR 5.5 vs DS 4.0) and fibula models (7-point Likert scale, VR 5.8 vs DS 5.1; Table 2). For the os coxae, the mean Hausdorff distance between VR and DS was 0.43 (SD 0.19) mm with a Dice coefficient of 96% (SD 2%). For the fibula, the mean Hausdorff distance between VR and DS was 0.29 (SD 0.12) mm with a Dice coefficient of 96% (SD 2%) (Table 1). The mean segmentation time of the os coxae models was 22.5 (SD 13.0) minutes for VR and 38.7 (SD 22.6) minutes for DS, and that of the fibula models was 12.1 (SD 8.0) minutes for VR and 17.0 (SD 11.1) minutes for DS (Figure 6; Table 1).

https://games.jmir.org/2023/1/e40541
Table 2. Results of the Likert-scale questionnaire after each of the 13 training cases.

<table>
<thead>
<tr>
<th>Likert (7-point) questions</th>
<th>Score in the VR(^a) environment (n=78), mean (SD)</th>
<th>Score in the DS(^b) environment (n=78), mean (SD)</th>
<th>Difference, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned something through this segmentation process.</td>
<td>5.7 (1.2)</td>
<td>4.9 (1.2)</td>
<td>0.8 (1.2)</td>
</tr>
<tr>
<td>I feel exhausted after the segmentation process.</td>
<td>4.4 (1.5)</td>
<td>5.2 (1.5)</td>
<td>−0.8 (1.8)</td>
</tr>
<tr>
<td>I had to repeat steps during segmentation.</td>
<td>4.3 (1.6)</td>
<td>5.0 (1.6)</td>
<td>−0.7 (1.6)</td>
</tr>
<tr>
<td>I consider this fibula to be adequately segmented.</td>
<td>5.8 (1.1)</td>
<td>5.1 (1.2)</td>
<td>0.7 (1.2)</td>
</tr>
<tr>
<td>I find this iliac crest to be adequately segmented.</td>
<td>5.5 (1.1)</td>
<td>4.0 (1.6)</td>
<td>1.5 (1.6)</td>
</tr>
<tr>
<td>I was able to orient myself spatially well during segmentation.</td>
<td>6.3 (0.7)</td>
<td>4.4 (1.3)</td>
<td>1.9 (1.7)</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.  
\(^b\)DS: desktop screen.

Figure 6. Comparison of segmentation time in minutes between the virtual reality (VR) environment (Elucis) in blue and the desktop screen (DS) environment (3D Slicer) in red shown as box-and-whisker plots. Box-and-whisker plots: boxes represent the IQR, thus representing 50% of data (Q1–Q3); the lower whisker is defined as Q1 − 1.5 × IQR; the upper whisker is defined as Q3 + 1.5 × IQR; the horizontal line in the middle of the box represents the median; and points are outliers. Additionally, the significance level of the respective Wilcoxon signed-rank test is shown above. The y-axis shows the segmentation time in minutes, and the x-axis shows the environment (VR vs DS). ***P<.001.

Training Curve
The linear mixed effect model (adjusted for work experience, mental rotation test, CT quality, segmentation quality, and training group) showed a significant increase in segmentation volume per minute (cm\(^3\)/min) for the VR environment (β=.86; P<.001). While the training effect for the DS environment was not significant anymore after \(P\) adjustment (β=.25; \(P=.26\); Figure 7). Lower CT quality had no influence on the segmentation process (\(P=.25\)). Segmentation performance was significantly higher in the VR environment than in the DS environment for the os coxae models (mean 19.2, SD 12.8 vs mean 12.0, SD 8.6 cm\(^3\)/min; paired \(t\) test, \(P<.001\)) and for the fibula models (mean 6.5, SD 2.9 vs mean 4.7, SD 4.0 cm\(^3\)/min; paired \(t\) test, \(P<.001\); Table 1). These results correlated with the results of the Likert-type questions concerning the learning effect, exhaustion, and the perceived need to repeat steps (Table 2).
Environment

Overall, the participants clearly preferred the VR environment over the DS environment in a poststudy questionnaire (Tables 3 and 4). The UEQ results showed that the participants assessed the VR environment to be better in terms of attractiveness, dependability, efficiency, novelty, perspicuity, and stimulation (Figure 8). The SUS results showed that Elucis was rated 83.3 (95% CI 75.3-90.8) and 3D Slicer was rated 30.4 (95% CI 20.1-38.3).

Table 3. Results of the Likert-scale questionnaire after completion of training.

<table>
<thead>
<tr>
<th>Likert (7-point) questions</th>
<th>Score in the VR&lt;sup&gt;a&lt;/sup&gt; environment (n=6), mean (SD)</th>
<th>Score in the DS&lt;sup&gt;b&lt;/sup&gt; environment (n=6), mean (SD)</th>
<th>Difference, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation in this environment was easy for me.</td>
<td>6.5 (0.5)</td>
<td>3.8 (1.9)</td>
<td>2.7 (2.0)</td>
</tr>
<tr>
<td>I would prefer this environment for VSP&lt;sup&gt;c&lt;/sup&gt;.</td>
<td>7.0 (0.0)</td>
<td>2.5 (0.5)</td>
<td>4.5 (0.5)</td>
</tr>
<tr>
<td>This work environment seems intuitive to me.</td>
<td>6.8 (0.4)</td>
<td>2.3 (1.0)</td>
<td>4.5 (1.0)</td>
</tr>
<tr>
<td>I suspect this work environment will continue to be the gold standard in clinical practice for VSP.</td>
<td>6.7 (0.8)</td>
<td>2.2 (0.8)</td>
<td>4.5 (1.4)</td>
</tr>
<tr>
<td>Monoscopic (DS) or stereoscopic (VR) vision in this environment made segmentation easy for me.</td>
<td>6.7 (0.5)</td>
<td>2.2 (1.2)</td>
<td>4.5 (1.5)</td>
</tr>
<tr>
<td>I prefer learning segmentation with a mouse and keyboard (DS) or VR controller (VR).</td>
<td>6.7 (0.5)</td>
<td>2.3 (0.8)</td>
<td>4.3 (1.2)</td>
</tr>
<tr>
<td>I found learning segmentation easy in this environment.</td>
<td>6.3 (0.5)</td>
<td>3.2 (1.2)</td>
<td>3.2 (1.2)</td>
</tr>
<tr>
<td>I felt the training scheme was appropriate for learning in this environment.</td>
<td>6.2 (0.8)</td>
<td>5.7 (1.5)</td>
<td>0.5 (1.5)</td>
</tr>
<tr>
<td>I felt the number of cases to learn segmentation was sufficient in this environment.</td>
<td>6.3 (0.8)</td>
<td>5.3 (1.2)</td>
<td>1.0 (1.7)</td>
</tr>
<tr>
<td>I had fun while learning segmentation in this environment.</td>
<td>7.0 (0.0)</td>
<td>3.8 (1.9)</td>
<td>3.2 (1.9)</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.

<sup>b</sup>DS: desktop screen.

<sup>c</sup>VSP: virtual surgical planning.
Table 4. Comments of the participants comparing the virtual reality and desktop screen environments.

<table>
<thead>
<tr>
<th>Participant No.</th>
<th>What aspects of the VR&lt;sup&gt;a&lt;/sup&gt; environment do you prefer over the DS&lt;sup&gt;b&lt;/sup&gt; environment?</th>
<th>What aspects of the DS environment do you prefer over the VR environment?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spatial vision, working with controllers, similarity to OP site.</td>
<td>Plausibility (you don’t have to take off the HMD), work more precisely with the mouse at times, other applications usable.</td>
</tr>
<tr>
<td>2</td>
<td>Similarity of actions to surgical activity, stereoscopic vision, good spatial and anatomical orientation.</td>
<td>Island tool.</td>
</tr>
<tr>
<td>3</td>
<td>3D perception, rotating and handling the dataset in your own hands, direct segmentation in rendered state; you can directly see the result of segmentation without segmenting in 2D layers.</td>
<td>The possibility of labeling each voxel separately in each layer, if desired; thus, higher accuracy is possible.</td>
</tr>
<tr>
<td>5</td>
<td>Very intuitive in comparison, optimal spatial orientation, orientation also with regard to the future operation area, faster learning.</td>
<td>It was possible to use “ADD” function in difficult cases (e.g. calcified vessels in direct neighborhood); calcified vessels and bones often have similar Hounsfield units and cannot be visualized separately in Elucis.</td>
</tr>
<tr>
<td>6</td>
<td>Easier to learn, 3D environment easier to control, better spatial overview, much better control.</td>
<td>Lower acquisition costs.</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.  
<sup>b</sup>DS: desktop screen.

Figure 8. User experience questionnaire (UEQ) comparison between the virtual reality (VR) environment (Elucis) in blue and the desktop screen (DS) environment (3D Slicer) in red shown as box-and-whisker plots. Box-and-whisker plots: boxes represent the IQR, thus representing 50% of data (Q1-Q3); the lower whisker is defined as Q1 – 1.5 × IQR; the upper whisker is defined as Q3 + 1.5 × IQR; the horizontal line in the middle of the box represents the median; and points are outliers. The Y-axis shows the UEQ score, and the X-axis shows the environment (VR vs DS). A total of 6 categories were examined (attractiveness, dependability, effectiveness, novelty, perspicuity, and stimulation).

Discussion

VSP in a traditional DS environment is still considered the gold standard, despite its disadvantages, such as being tedious and time consuming [43]. This negates the benefit of time saved in the operating theater due to additional time spent on VSP [13,14]. In this context, VR is a promising approach that might
attenuate the mentioned disadvantages of VSP performed in a DS environment.

In this study, the processes of learning VSP, using segmentation as a surrogate parameter, between traditional DS and fully immersive VR environments were compared, as segmentation is a common substep in VSP, a bottleneck in terms of time [19,44], and a crucial factor that determines the accuracy of the final computer-aided design (CAD) result [45]. In general, segmentation is important for not only oral and maxillofacial surgery but also many other medical specialties that require 3D models created by segmentation for the subsequent steps in CAD [45]. Therefore, segmentation was used as a surrogate parameter to measure performance in VR compared to performance in a traditional DS environment during a structured training course. For a representative number of cases (n=156), all participants had to create models of the fibula and os coxae, which are commonly used for microvascular reconstruction of the mandible [2,3], in both environments.

Many studies have shown that VSP is beneficial to patients [14,46]. However, VSP is only feasible if the surgeon is proficient in VSP software. For this reason, the process of learning how to operate such CAS-related software has already been the subject of research in the literature. For example, learning curves upon repetition have been found for VSP for the treatment of orbital fractures or in orthognathic surgery [15,47]. In addition, differences in learning curves when it comes to CAD programs used in dentistry have been found between different software and professional groups (students, technicians, and dentists) [18,33]. Furthermore, the learning curve has been shown to be a function of the complexity of the task for model creation using CAD [48]. However, all these studies used applications in a 2D DS environment, where a computer mouse and keyboard were utilized. This type of work environment is counterintuitive to surgeons who are accustomed to ambidextrous, manual, and visuospatial activity while being very focused. In contrast, VR is more similar to surgical work and therefore offers a promising approach.

The results of our study showed that working in a VR environment is significantly more efficient than working in a DS environment. At the same time, VR has a much faster learning curve (Figure 7). To have a comparable measure between models with different volumes, we used segmentation volume per minute (cm³/min), assuming that it significantly reflects the number of attempts to solve the task during the acquisition of cognitive and motor skills [48]. At the end of the training program, this was almost twice as high in the VR environment compared to the DS environment. To account for a transfer effect from one method to the other due to the crossover study design, we split the groups in correspondence to similar studies [18,33] and adjusted our model for training method order. To exclude bias from the multiple steps of the VSP as described in other studies that have evaluated the learning process for an entire VSP/CAS based on the final surgical outcome [15,49], we focused on the process performance itself. The measurable stronger learning effect in the VR environment was also perceived by the participants (Table 2). This is important, as the faster learning of CAD software, such as that for VSP, is also accompanied by increased planning and surgical accuracy, and improved anatomical understanding of the target structures [15,47]. Independent of segmentation performance, it is assumed that for different DS software, with correspondingly longer training periods, the differences will eventually cease to exist [33]. However, this does not explain the large differences observed between the DS and VR environments.

These differences can be explained by the principally different natures of the DS and VR environments. Subjects in our study reported a high degree of anatomical orientation in the VR environment compared to the DS environment (P<.001). One reason for the superior anatomical orientation might be the possibility of obtaining different viewing angles in the VR environment. Specifically, Cha et al demonstrated this advantage [24]. Furthermore, in our study, subjects preferred to work with both hands using 2 VR controllers, appreciated the advantages of stereoscopic vision, and preferred the intuitive perception provided by the VR environment, which might explain their preference for the VR environment over the DS environment overall (Table 3). In summary, this study highlighted the benefits of spatial perception and working in 3D in a VR environment with good orientation in terms of the anatomy and strong similarity to surgical work. Similar results have been found in the use of nonfully immersive VR for CAS of the liver [21]. In the literature, various approaches to using at least nonfully immersive VR in the medical context can be found. For example, stereoscopic monitors have been used in oral and maxillofacial surgery to determine facial soft tissue or orbital volumes [50,51]. Furthermore, they have been used in neurosurgery to train surgeons to resect intracerebral tumors [52] and assess surgical risk as a rehearsal before surgeries [25,53].

Yet, all of the aforementioned approaches lacked the full immersion that comes with using VR HMDs and controllers in both hands, as in our study. In our setting, users held control elements in both hands and worked with them within arm’s reach (Figures 2 and 3). This could be another reason for the rapid learning and efficiency of working in VR, as proprioception can then be used. Therefore, subjects felt particularly comfortable, and the highest precision could be achieved [54]. In contrast, working with a computer mouse and keyboard in the DS environment does not allow this. For complex models (os coxae) in comparison to simpler models (fibula), this advantage was even more pronounced in VR than in DS. In addition to the advantages of the stereoscopic visualization of complex models in VR [55], the reasons for this are that complex models require more commands for their implementation [48] and, at the same time, bimanual work in VR is particularly faster [56].

Regarding the segmentation quality of the os coxae model, a certain discrepancy between the radiologists’ evaluations and that of the subjects appeared to be a shortcoming, as the subjects tended to overestimate themselves. However, this had no impact on clinical practice, since with a given clinically relevant score of ≤2 (good or excellent), we found no significant difference between the 2 environments. A possible reason for this overestimation could be that in the VR environment by default,
usually only 1 slice (axial, coronal, or sagittal) was displayed. In contrast, all 3 slices were displayed in parallel in the DS application. One solution would be to display all 3 slices by default in VR. In addition, a different perception of the 3D models (stereoscopic during segmentation by participants vs monoscopic during evaluation by radiologists) could have influenced this. Another reason could be the resolution capacity of the used VR HMD, causing individual voxels to be blurred. Improved hardware or better delineation of individual voxels in the single-layer view could probably solve this.

Overall, our results suggest that the intrinsic feedback learning effect of VSP, which a surgeon achieves by multiple repetitions of VSP and consecutive surgeries [15], could potentially be at least partially replaced by a training program in this manner for novices in VSP. Therefore, the results of our study suggest that VR should be the environment of choice, given its similarity to actual surgical activity. Furthermore, subjects consistently reported less fatigue and more enjoyment while working in the VR environment (Tables 2 and 3). This was also in line with the SUS results, which clearly rated VR better than DS. The reduced exhaustion can be explained by the shorter time needed to complete the task and possibly by the manual 3D interactions in VR, resembling the surgical working environment more realistically (Table 2). Participants perceived the VR software to be superior to the DS software according to the UEQ results for multiple qualities of experience (Figure 8) and preferred the VR environment over the DS environment (Table 3). This was in line with the results of a study on VR user interfaces for medical marking on 3D models [57]. Yet, these differences are surprising, since both applications basically fulfill the same task.

Experiencing the burden of today’s required levels of documentation, surgeons have long been demanding to spend less time at the computer [58]. This could probably also explain the poor rating for the DS environment, which is similar to the typical unpopular work on a DS computer. However, VR has a more playful character, which might explain the higher ratings for the VR environment. However, this clear difference in rating also shows that VSP applications should be adapted to the needs of surgeons. VR seems to meet these requirements, at least in the case of our test persons. In this context, the following things should be considered. Using a VR environment can lead to VR sickness. However, this was not reported by any of the participants, probably because they were sitting and no dynamic scenes were used [59]. In addition, it should be considered that older adults take longer to adapt when learning new technologies [60]. In this regard, studies have shown that older adults also adapt well to VR [61]. Unlike DS with the use of a keyboard and mouse, in the performance of VSP by older adults, VR can be advantageous, as VR is closer to actual surgery due to the immersive environment, stereoscopic vision, and bimanual working. Another point that needs to be considered is cost. Such a system would have to be acquired first. On the other hand, the use of VR, as shown in our study, is time-saving and could thereby reduce the overall cost of VSP. However, an investigation of cost was outside the scope of our study.

Trials featuring the completion of VSP in a nonfully immersive manner have been described to be very promising in the field of liver surgery [21]. However, to our knowledge, no studies have yet assessed VSP performed completely in fully immersive VR. Future studies should address, aside from segmentation, the other steps of full VSP, with the goal to perform VSP completely in VR.

Apart from the approach shown in our study, semiautomated or fully-automated segmentation algorithms are increasingly being used to accelerate or completely take over segmentation [20]. Yet, our study is concerned about which is the better working environment to perform any kind of VSP. This is even more important since the most common segmentation algorithms require a high amount of time for manual postprocessing [45], and the sole use of fully automated algorithms has still not gained general acceptance in clinical practice due to their high sensitivity to image-related artifacts, a resulting inadequate structure recognition, and a low functional stability [20]. In the future, however, artificial intelligence and VR could be combined, especially to further accelerate tasks in VR.

We not only demonstrated the feasibility of segmentation in VR but also highlighted the advantages of this. Possible explanations for these findings are the more intuitive work interface provided by the VR environment, the better spatial orientation, and the stereoscopic vision [21], which were perceived as great advantages for the purposes of VSP.

Through the crossover design, the subjects represented their own control group, and with a total of 312 experiments, the study achieved good internal validity. To improve the generalization of the results, that is, the external validity, we performed the study with subjects having different surgical experience. However, a greater diversity of subjects could lead to different results. The external validity should be evaluated in a further study, and a multicenter study is preferred to include subjects with experience from other clinics. However, this does not diminish the significance of this study. As with the various phases of drug studies, it is worthwhile to first examine the internal validity. Based on the findings of this study, a subsequent study for good external validity can be planned with more subjects but fewer examinations per subject. Furthermore, it should be considered that 2 different software programs (3D Slicer and Elucis) were compared with each other. Although both programs have a comparable command set, a slight impact cannot be excluded. However, this would not explain the large differences observed.

In contrast to the DS environment, the VR environment offers a more rapid learning curve, provides the ability to work faster, results in increased time saved, and is less exhausting when engaging in segmentation as an important substep of VSP. The VR environment was rated by participants to be easier to learn and to be their environment of choice, possibly due to its similarity to actual surgical activities. Therefore, VR-based VSP could be better integrated into clinical practice. This study can serve as a basis for investigating other common substeps of VSP.
Acknowledgments
We thank Rainer Röhrig for his valuable feedback and suggestions. We would like to thank Justin Sutherland and Sonya Deveau from Realize Medical Inc for providing us with a free license of Elucis to be able to perform this study.

Data Availability
The data presented in this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
BP, MU, AB, FH, and TWK conceptualized the study; MU, BP, and VV contributed to the methodology; MU, BP, VV, CW, and AR contributed to validation; MU, BP, AB, and LJG performed formal analysis; MU, BP, LJG, MO, CM, and IM performed the investigation; BP, FH, and TWK managed the resources; MU, BP, and VV contributed to data curation; MU wrote the original draft; BP, MU, MO, VV, AB, LJG, CM, IM, CW, AR, TWK, and FH reviewed and edited the manuscript; MU, BP, and MO contributed to visualization; BP and FH contributed to supervision; and BP managed project administration. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References


42. Dikta G, Scheer M. Bootstrap Methods With Applications in R. Cham: Springer; 2021.


**Abbreviations**

- **CAD**: computer-aided design
- **CAS**: computer-assisted surgery
- **CT**: computed tomography
- **DS**: desktop screen
- **HMD**: head-mounted display
- **SUS**: System Usability Scale
- **UEQ**: User Experience Questionnaire
- **VR**: virtual reality
- **VSP**: virtual surgical planning

©Max Ulbrich, Vincent Van den Bosch, Andrea Bönsch, Lennart Johannes Gruber, Mark Ooms, Claire Melchior, Ila Motmaen, Caroline Wilpert, Ashkan Rashad, Torsten Wolfgang Kuhlen, Frank Hölzle, Behrus Puladi. Originally published in JMIR Serious Games (https://games.jmir.org), 19.01.2023. 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.
Design and Evaluation of Using Head-Mounted Virtual Reality for Learning Clinical Procedures: Mixed Methods Study

Siew Tiang Lau¹, BHS, MHS, PhD; Rosalind Chiew Jiat Siah¹, BHS, MHS, PhD; Khairul Dzakirin Bin Rusli¹, BSc; Wen Liang Loh¹, BA, Grad Dip (Psych); John Yin Gwee Yap¹, BA, MA; Emily Ang¹, DNurs; Fui Ping Lim¹, BHSN, MNurs; Sok Ying Liaw¹, BHS, MHS, PhD

Alice Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Siew Tiang Lau, BHS, MHS, PhD
Alice Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 2, Clinical Research Centre, Block MD 11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011299
Email: nurlst@nus.edu.sg

Abstract

Background: The capacity of health care professionals to perform clinical procedures safely and competently is crucial as it will directly impact patients’ outcomes. Given the ability of head-mounted virtual reality to simulate the authentic clinical environment, this platform should be suitable for nurses to refine their clinical skills for knowledge and skills acquisition. However, research on head-mounted virtual reality in learning clinical procedures is limited.

Objective: The objectives of this study were (1) to describe the design of a head-mounted virtual reality system and evaluate it for education on clinical procedures for nursing students and (2) to explore the experience of nursing students using head-mounted virtual reality for learning clinical procedures and the usability of the system.

Methods: This usability study used a mixed method approach. The stages included developing 3D models of the necessary instruments and materials used in intravenous therapy and subcutaneous injection procedures performed by nurses, followed by developing the procedures using the Unreal Engine (Epic Games). Questionnaires on the perception of continuance intention and the System Usability Scale were used along with open-ended questions.

Results: Twenty-nine nursing students took part in this questionnaire study after experiencing the immersive virtual reality (IVR) intervention. Participants reported largely favorable game perception and learning experience. Mean perception scores ranged from 3.21 to 4.38 of a maximum score of 5, while the mean system usability score was 53.53 of 100. The majority found that the IVR experience was engaging, and they were immersed in the game. The challenges encountered included unfamiliarity with the new learning format; technological constraints, such as using hand controllers; and physical discomfort.

Conclusions: The conception of IVR for learning clinical procedures through deliberate practice to enhance nurses’ knowledge and skills is promising. However, refinement of the prototypes is required to improve user experience and learning. Future research can explore other ways to use IVR for better education and health care purposes.

(JMIR Serious Games 2023;11:e46398) doi:10.2196/46398

KEYWORDS
user experience; acceptability; usability; virtual patient; clinical procedure; immersive; nursing student; virtual reality; education; performance
Introduction

Overview

Generation Z students are well recognized to have different learning needs and preferences than previous generations [1]. With the advancement of medical technology and treatment, coupled with dynamic clinical settings, the ability of nurses to perform clinical procedures safely and competently is of paramount importance. The issue of ensuring competency prior to performance on actual patients remains crucial for patient safety. Therefore, new teaching strategies are needed to bridge the teaching-learning gap to create an effective learning environment. The use of technology to meet the challenges of medical education has been well documented [2,3]. Simulation sessions using manikins in clinical laboratory settings have been established for more practice. However, these simulation sessions are usually conducted in small groups and are labor and resource intensive. Studies also suggest that these conventional teaching methods do not fully meet the needs, interests, and learning preferences of Generation Z students [4,5]. The novel discovery of virtual technology has been explored for its effectiveness in enhancing students’ learning in nursing education [3,6,7].

Immersive virtual reality (IVR) is highly visual and conducive to procedure-focused learning. IVR provides first-person active learning where learners can interact with 3D objects and characters [8]. With immersive 3D virtual reality (VR), learners can perform clinical procedures in a real clinical environment. It is well established that IVR as a medium of instruction facilitates interaction and immersion [9]. This enhances learning through the learner’s psychological constructs, such as presence and agency. Presence forms the conditions, and agency gives autonomy of control to gain experience through immersive technology [9]. In designing IVR training, it is important to understand the factors influencing learning. Factors such as perceived realism, usability, and learning are crucial to learners’ attitude toward the use of VR [10]. Hence, VR features, learning experiences, usability, and learning theories that guide the learning process are vital constructs when designing an IVR training environment [10].

The transformation of nursing education has been rapidly evolving with the advancement of VR technology. IVR allows learners to obtain repeated hands-on practice to understand concepts and master procedures, exercise their problem-solving skills, and develop confidence prior to managing actual patients in clinical settings. IVR allows learners to practice safely without risk to patients and empowers them for self-directed and regulated learning. IVR was reported to be used frequently to teach procedural-practical knowledge [11]. Although there have been nursing VR simulations conducted to improve technical knowledge and proficiencies, these studies focused mainly on linear step-by-step intervention, and the order of the steps played an important role [12,13]. Systematic reviews evaluating the effectiveness of IVR report that it improved knowledge and learning experiences, but the reviews’ findings were inconclusive for skills, satisfaction, confidence, and performance in health care education [14-17].

Theoretical Framework

The conceptual frameworks that guided this study are cognitivism, constructivism, and experiential learning through deliberate practice [18]. Cognitivism theory has been used for information processing in the context of VR. Learning in VR involves immersion, interaction, imagination, motivation, and transformation of new knowledge to enhance problem-solving capability [18,19]. Through the constructivism approach, students engage with the environment and are actively involved in building a scaffold for their learning to construct and develop new skills and meaningful knowledge [20,21]. Experiential learning refers to knowledge creation through the acquisition and transformation of experiences [22]. Through first-hand individual experiences using IVR, students are motivated to learn more in a safe and convenient environment for practice. The interaction with a task in VR allows the construction of learning experiences via abstract, active, and reflective processes of the experiential learning cycle [21,23]. With experiential learning and timely built-in feedback, students will be able to work on areas that they need to improve through self-analysis and reflection and achieve competence through meaningful learning.

Deliberate practice is the repetitive performance of a cognitive or psychomotor task in a focused domain combined with rigorous skills assessment [24]. Through deliberate practice, students learn new knowledge, analyze, reflect, and problem-solve cyclically for self-improvement [22,25]. With mastery of their experience, students can use deliberate practice, receive feedback on their own performance, improve self-efficacy, and construct new knowledge through experience, as shown in Figure 1.

The aims of this study were (1) to describe the design of a head-mounted virtual reality system and evaluate it for education on clinical procedures for nursing students and (2) to explore the experience of nursing students with head-mounted VR for learning clinical procedures and the usability of the system.
Methods

Design and Implementation of IVR Clinical Procedures

Overview

In this study, 4 key elements were considered when designing the VR prototype to create an engaging and immersive experience for users. The first element was accessibility and inclusivity for all learners and involved software components, adjustable height, and visual and audio design to minimize cybersickness. The second element was locomotion and navigation to allow the learners to navigate the virtual space intuitively. The third element was visual fidelity; the graphics were designed in a detailed and realistic way to provide good immersion for the IVR experience. The fourth element was motion tracking, where the learner’s head and hand movements were detected to allow real-time interaction. An additional consideration was the degrees of freedom and interactivity whereby the learner was given control of their learning [9].

A real-time 3D creation tool, Unreal Engine 5 (Epic Games), was used to support the development of the IVR procedures, which reinforced the real-time interaction of the learner, clinical instructor, and patient in a virtual ward [26]. The design and development of the IVR clinical procedures were a close collaborative effort among the nursing school, health care institution, and IT experts in a systematic process. First, the faculty and nurse clinicians discussed and created clinical scenarios and consolidated common errors, critical steps, and safety issues pertinent to each procedure during the developmental stage.

Then, students accessed the proposed immersive simulation of the clinical procedures by wearing a 3D VR headset with 2 handheld controllers (Oculus Quest 2; Reality Labs), shown in Figure 2. Using the Oculus Quest, each student could visually experience a typical clinical session modeled after a real laboratory with its procedures, which they might have experienced before in their typical laboratories. The simulation was highly interactive, leveraging the handheld controllers for the left and right hands to enable and engage the students in executing the same simulated procedures vividly seen in their facilitated laboratory demonstrations. With the rich visual media and interactive experience, students could practice the simulated virtual procedures in their own time and space without mental stress, external distractions, fear of hazardous consequences, or creation of any clinical waste.

Within the IVR program, 3 stages for playing using the scaffold learning approach, namely, the orientation, practice, and assessment modes, were developed.

Figure 1. Conceptual framework for learning clinical procedure using virtual reality.
Orientation

New users were guided by a tutor and an instructional video on the features and safety aspects of using the IVR clinical procedures for learning. When they put on the Oculus Quest 2 for the first time, they were required to complete virtual tutorials on how to maneuver the controllers, how to perform handwashing by rubbing 2 hands together using the controllers (Figure 3) and putting on and removing gloves to optimize their gaming experience.

Practice Mode

Visual cues were provided when the learners learned how to proceed from one step to the next. With the built-in checklist, ample opportunities were given for the learners to learn from their experiences. With timely feedback based on the student’s responses, they could rationalize their actions and reactions and thus decide the next course of action or repeat the session. Finally, after the learners gained confidence through practice, made sense of the procedures, and built on their prior learning, they were ready for assessment. No cues were given, and the students could see their results and repeat the practice when necessary.

Figure 2. Oculus Quest (with 2 hand-held controllers).

Figure 3. Hand hygiene.
required. This approach aimed to allow the construction of knowledge through experience and deliberate practice. It boosted self-regulated learning, as learners could adjust their learning according to their own pace to either move to the next level or remain in the practice mode until they were ready for assessment [27].

Assessment Mode
No cues were provided, as this was a simulation of an actual clinical situation. The learner was expected to proceed according to what they had already learned and practiced. Reward systems were incorporated, with a maximum of 3 stars when the learner completed the procedures correctly. Deliberate practice and timely feedback on the learner’s performance were integrated into the case scenarios and situations to develop critical thinking and problem-solving abilities. Each IVR experience lasted 15-20 minutes to ensure no learner fatigue.

Implementation
Figures 4 and 5 illustrate the subcutaneous injection and intravenous therapy procedures. In each scenario, the learners performed the procedures using the 2 handheld controllers. They also interacted with the patient and clinical instructor during critical steps such as checking identifiers and the “5 rights” when administering medication to ensure patient safety. The learners could click on a text box, and the patient would reply verbally with prerecorded responses. For example, if the learner asked, “May I know what your name and NRIC number are?” the virtual patient would reply, “My name is Ahmad, NRIC number is S1234567x.” An electronic scanner and medication record were created to provide realism for learners to check the patient and medication prior to the administration.

The virtual simulation of administering a subcutaneous injection (Figure 4) included the critical component of checking the correct instruments and materials, type of insulin, and techniques in withdrawing air prior to withdrawing the correct unit from the vial. The appropriate injection site and any complications were identified, such as pain, edema, erythema, temperature, and lipohypertrophy. Critical components, such as checking patient identifiers, hand hygiene, the correct 5 rights, and injection techniques, had to be performed correctly to pass the procedure.

For the intravenous therapy procedure (Figure 5), learners studied critical concepts such as using the correct fluids and the accurate rate of infusion when administering intravenous therapy. Safety aspects, such as maintaining aseptic techniques while spiking the drip set and connecting to the intravenous plug, and steps to prevent phlebitis were incorporated into the scenarios. The learners were also expected to examine the risk of air embolism by checking the intravenous tubing.

Figure 4. Subcutaneous injection.
Evaluation of IVR Clinical Procedure

Study Design
This usability study used a mixed methods approach using convenience sampling. The recruitment and intervention of participants took place from June to August 2021 upon receiving approval from the institutional ethics review board.

Participants
The participants comprised undergraduate nursing students in their second academic year who had completed the pathophysiology and pharmacology modules and the Nursing Practice 1 and Clinical Experience Practicum 1.2 modules.

Data Collection
To achieve the aims of the study, three questionnaires were administered to the participants, which collected the following information: (1) basic demographics and past VR experience, (2) perception of continuance intention, with an instrument developed by Roca et al [28], and (3) the System Usability Scale (SUS) developed by Brooke [29]. To maintain consistency, we explained how to use the head-mounted device and how to put it on to ensure it was in good working condition. Any device issues were rectified, and then the participants brought the IVR device home for 2 weeks of testing on their own time. Upon returning the device, they were invited to complete the questionnaires.

The scale for perception of intention to continue was adapted from Roca et al [28] and had the following subscales: (1) usefulness, (2) cognitive absorption, (3) ease of use, (4) system quality, (5) confirmation, (6) satisfaction, and (7) continuance intention. All items were rated on a 5-point scale (1=strongly disagree to 5=strongly agree). The scores for each item were summed and multiplied by 2.5 to calculate a total score ranging from 0 (lowest) to 100 (highest). There was no cutoff score, but higher total scores indicated higher levels of usability for the IVR procedures. The Cronbach α values for the scale ranged from .85 to .91 [30].

Quantitative Analysis
Descriptive statistics using percentages, means, and SDs were computed to examine the study variables using SPSS (version 28.0; IBM Corp) [31].

Qualitative Analysis
Qualitative data from the open-ended questions were imported into an Excel spreadsheet for coding. Content analyses were used to understand the users’ experiences of the IVR procedure [32]. Two researchers (STL and WLL) conducted the content analysis and extracted the codes, categories, and themes [33]. The research team reviewed the frequencies of the coded themes to reach a consensus. Trustworthiness, credibility, and transferability were ensured by including all 29 participants’ responses, selecting units of meaning, and presenting the findings in a rich and vigorous way (Multimedia Appendix 1).

Ethics Approval
Ethical approval was obtained from the National University of Singapore Institutional Review Board (NUS-IRB 2021-305).
The participants were briefed thoroughly, and the participant information sheet was provided prior to recruitment. Participants were informed that their participation was voluntary and that they could withdraw from the study at any time. Confidentiality and anonymity were ensured.

**Results**

**Participants**

In this study, 29 participants were recruited, and all completed the study. The mean age of the participants was 23.03 (SD 6.59) years, with the majority being female (n=18, 62%). There were 22 Chinese participants (76%), 3 Malay participants (10%), and 1 Indian (3%) participant, with 3 participants of other ethnicities (10%). Most had no experience using VR prior to the user testing (n=27, 93%).

**Quantitative Analysis**

As presented in Table 1, for perception on intention, all the subscales in the perception scale had a mean score greater than 3 (of a total score of 5). This indicated that the students had a positive perception of the quality and usability of IVR, which is a key determinant for their continuity intention. The subscale with the highest mean score was “cognitive absorption” (mean 4.05, SD 0.61), suggesting that the students enjoyed and found the IVR engaging. The top 2 items with the highest scores supported this notion: “I have fun interacting with VR procedure” (mean 4.38, SD 0.68) and “When I am using the VR procedure, I am absorbed in what I am doing” (mean 4.14, SD 0.74). From another aspect, the 2 items with the lowest scores were “It is easy for me to become skillful at using VR procedure” (mean 3.21, SD 1.01) and “I will frequently use the VR procedure in future” (mean 3.24, SD 0.95).

However, for the student’s perception of the usability of IVR, the mean SUS score was 53.53 (SD 16.18). This score was below the acceptable score of 68, indicating that more work needed to be done to improve the usability of IVR. Two items with poor perception, that was, items with >10 participants, indicating “Do not agree for positively worded item” or “Agree for a negatively worded item,” were “I thought the VR procedure was easy to use” and “I think that I would need assistance to be able to use the VR procedure” (Table 2).
Table 1. Perception scale (N=29).

<table>
<thead>
<tr>
<th>Subscales or items</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Usefulness</strong></td>
<td></td>
</tr>
<tr>
<td>Using the VR(^a) procedure can improve my learning performance</td>
<td>3.79 (0.90)</td>
</tr>
<tr>
<td>Using the VR procedure can increase my learning effectiveness</td>
<td>3.66 (0.97)</td>
</tr>
<tr>
<td>I find the VR procedure to be useful to me</td>
<td>3.69 (1.11)</td>
</tr>
<tr>
<td><strong>Cognitive absorption</strong></td>
<td></td>
</tr>
<tr>
<td>Time flies when I am using the VR procedure</td>
<td>4.00 (0.96)</td>
</tr>
<tr>
<td>Most times when I get on to the VR procedure, I end up spending more time than I had planned</td>
<td>3.69 (0.81)</td>
</tr>
<tr>
<td>When I am using the VR procedure, I am absorbed in what I am doing</td>
<td>4.14 (0.74)</td>
</tr>
<tr>
<td>I have fun interacting with VR procedure</td>
<td>4.38 (0.68)</td>
</tr>
<tr>
<td>I enjoy using the VR procedure</td>
<td>4.03 (0.82)</td>
</tr>
<tr>
<td><strong>Ease of use</strong></td>
<td></td>
</tr>
<tr>
<td>Learning to operate the VR procedure is easy for me</td>
<td>3.31 (1.04)</td>
</tr>
<tr>
<td>It is easy for me to become skillful at using VR procedure</td>
<td>3.21 (1.01)</td>
</tr>
<tr>
<td>My interaction with the VR procedure is clear and understandable</td>
<td>3.45 (0.91)</td>
</tr>
<tr>
<td><strong>System quality</strong></td>
<td></td>
</tr>
<tr>
<td>Number of steps per task is too many(^b)</td>
<td>3.55 (0.63)</td>
</tr>
<tr>
<td>Steps to complete the task follow a logical sequence</td>
<td>3.62 (0.86)</td>
</tr>
<tr>
<td>The organization of information is clear</td>
<td>3.41 (0.98)</td>
</tr>
<tr>
<td>The VR procedure has natural and predictable screen changes</td>
<td>3.55 (0.91)</td>
</tr>
<tr>
<td>The VR procedure system is responsive</td>
<td>3.34 (1.01)</td>
</tr>
<tr>
<td><strong>Confirmation</strong></td>
<td></td>
</tr>
<tr>
<td>My experience with using the VR procedure was better than I expected</td>
<td>3.48 (1.12)</td>
</tr>
<tr>
<td>The service level provided by the VR procedure was better than I expected</td>
<td>3.55 (0.87)</td>
</tr>
<tr>
<td>Overall, most of my expectations from using the VR procedure were confirmed</td>
<td>3.52 (0.78)</td>
</tr>
<tr>
<td><strong>Satisfaction</strong></td>
<td></td>
</tr>
<tr>
<td>I am satisfied with the performance of VR procedure</td>
<td>3.31 (1.04)</td>
</tr>
<tr>
<td>I am pleased with the experience of using the VR procedure</td>
<td>3.83 (0.89)</td>
</tr>
<tr>
<td>My decision to use the VR procedure was a wise one</td>
<td>3.97 (0.98)</td>
</tr>
<tr>
<td><strong>Continuance intention</strong></td>
<td></td>
</tr>
<tr>
<td>I will use the VR procedure on a regular basis in future</td>
<td>3.34 (1.08)</td>
</tr>
<tr>
<td>I will frequently use the VR procedure in future</td>
<td>3.24 (0.95)</td>
</tr>
<tr>
<td>I will strongly recommend others to use it</td>
<td>3.79 (1.11)</td>
</tr>
<tr>
<td>Overall score</td>
<td>3.63 (0.63)</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.

\(^b\)Scored in reverse.
Table 2. System usability scale (N=29). The total score was 53.53 (SD 16.18).

<table>
<thead>
<tr>
<th>Items</th>
<th>Strongly agree or agree, n (%)</th>
<th>Neutral, n (%)</th>
<th>Strongly disagree or disagree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positively worded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think I would like to use VR procedure frequently</td>
<td>14 (48)</td>
<td>10 (34)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>I thought the VR procedure was easy to use</td>
<td>10 (34)</td>
<td>8 (28)</td>
<td>11 (38)</td>
</tr>
<tr>
<td>I found the various functions of VR procedure were well integrated</td>
<td>12 (41)</td>
<td>10 (34)</td>
<td>7 (24)</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use VR procedure very quickly</td>
<td>11 (38)</td>
<td>10 (34)</td>
<td>8 (28)</td>
</tr>
<tr>
<td>I felt very confident using the VR procedure</td>
<td>8 (28)</td>
<td>12 (41)</td>
<td>9 (31)</td>
</tr>
<tr>
<td>Negatively worded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I found the VR procedure unnecessarily complex</td>
<td>7 (24)</td>
<td>10 (34)</td>
<td>12 (41)</td>
</tr>
<tr>
<td>I think that I would need assistance to be able to use the VR procedure</td>
<td>13 (45)</td>
<td>6 (21)</td>
<td>10 (34)</td>
</tr>
<tr>
<td>I thought there were too much inconsistency on the VR procedure</td>
<td>6 (21)</td>
<td>7 (24)</td>
<td>16 (55)</td>
</tr>
<tr>
<td>I found VR procedure very cumbersome or awkward to use</td>
<td>9 (31)</td>
<td>8 (28)</td>
<td>12 (41)</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with the VR procedure</td>
<td>8 (28)</td>
<td>9 (31)</td>
<td>12 (41)</td>
</tr>
</tbody>
</table>

aVR: virtual reality.

Qualitative Analysis

Overview

Key themes were derived from the content analysis of the open-ended questions on the questionnaires regarding the experience of using VR procedures for learning. The themes derived from the analyses included the usefulness of IVR, learning experiences using IVR, challenges encountered, technical glitches, and suggestions for improvement. Themes and frequencies from the content analysis can be found in Multimedia Appendix 1.

Usefulness of IVR

Participants were excited about the new experience of using IVR. They found that the realism and relevance to the clinical setting and the ability to interact with a virtual patient and clinical environment promoted immersion in experiential learning:

*It is intriguing as a first-time user; hence it caught my attention from the start. I could practice the nursing procedures at my own comfortable pace and get back to practicing the procedures again on my own time. There was an improvement in terms of accuracy in the procedure as I practiced more often on my own. Having a virtual patient in front of our eyes helped mimic clinical settings.* [P11]

Participants appreciated this learning format and found it a useful resource for students to practice and learn at their own pace, time, and convenience. “I can see this being a very useful resource for students to practice and learn in their own time” [P15].

Learning Experiences

Participants reported that performing the procedure using IVR encouraged them to consciously revise the sequence for the procedure and ensure accuracy as they built on their knowledge and skill through practice. This experience enhanced their thinking skills as they reviewed their performance:

*I liked that the VR prompted me to think through the sequence of steps I have to do before a procedure. It helped me revise the steps and rationale behind the steps. I also like that the VR tracks your activity and lets you know which step you have forgotten to perform.* [P3]

Some participants commented that learning on the VR platform allowed novice learners to correct mistakes without putting an actual patient at risk:

*The fact that I can practice clinical skills anywhere is an advantage of VR. Also, I would not be causing any real patient harm if I tried something out for the first time.* [P23]

*It will make me understand my mistakes there, so I will be more aware not to repeat them the next try or during real-life clinical procedures.* [P11]

In particular, the participants appreciated the important concepts that might be overlooked in real clinical settings. One of these aspects was performing the “5 moments of hand hygiene” for any patient encounter. With IVR, the participants could not proceed, and they could see the “dirt” visually on their hands. “The visualization of soiled hands was also very useful as I see the importance of hand hygiene” [P21].

Challenges Encountered

Participants reported some limitations of the IVR system. The participants shared that adapting to the device and controls took time. Some concerns included physical discomfort, such as feeling dizzy, and that participants could not complete the entire procedure in 1 sitting:

*I could not complete the tasks because I felt dizzy after a certain amount of time, and though I do revise...*
the steps before I put on the headgear to do the task.  
[P27]

Some of them also commented on their unfamiliarity with the device and instructions. Moreover, others were unable to communicate with the virtual patient and had a decreased level of enjoyment in using IVR:

I feel like there can be clearer instructions, possibly step by step, during my first attempt of the procedure to refresh and familiarize myself with the items. I could not communicate with the patient, grasp objects, put on gloves, and check the patient’s body.  
[P18]

**Technical Glitches**

The most technical glitches involved the withdrawal and administration of the medication procedure. The participants shared that they experienced irritation when virtual items were not responding according to commands despite multiple attempts. This impeded their progress and learning as they needed to restart the procedures:

Some IT interface needs to be adjusted to smoothen the flow of procedural execution. Sometimes there is a glitch where the packaging of the alcohol swabs, needle, and IV line gets stuck to our hand, thus hindering our next steps.  
[P26]

It was very frustrating that I could not inject the needle into the insulin vial and manipulate it too.  
[P22]

**Suggestion for Improvement**

Three main areas for improvement were suggested: provide orientation and more specific instructions, include more audio and haptic feedback in the game, and optimize the controls and joystick. Some other suggestions included adding a progress bar to motivate students to practice, including more procedures, and extending the use to hospitals:

More instructions can be given. A good example would be IV priming, where additional instructions are taught to us.  
[P5]

Please include more procedures…look forward to it.  
If it works well, this can be a good technology to bring into hospitals for staff re-education or enrichment purposes. And any updates to a particular nursing skill can be updated in the software as well...that will be excellent.  
[P6]

**Discussion**

**Overview**

The objectives of this study were to describe the development and to evaluate the learning experience of undergraduate nursing students. Creating these 2 clinical procedures in VR involved multiple stakeholders, including multiple reviews and refinements prior to this study. A new teaching approach was incorporated to create an effective learning environment to promote competencies among nurses. This study was a timely development during the period of the pandemic, when there were movement restrictions, safe distancing, and physical segregation of staff and students. More essentially, this study met the learning needs of the next generation.

**Principal Findings**

This study indicated that undergraduate nursing students found IVR satisfying and most of them enjoyed the experience. They were most intrigued by the contextualization of the VR procedures and felt that the experience was better than they had expected. This result reflects findings in the broader literature [34]. Once in the VR environment, they were absorbed in performing the procedures and tended to spend more time than intended. One of the reasons could be the high degree of realism, as the VR environment was a close representation of the actual ward setting. The participants could interact with patients when guided by the clinical instructor, just like in real life. From the study, the scores in the perception scale and qualitative data showed that the participants were highly engaged and immersed and forgot the time when performing the VR procedures, which is congruent with other studies [35].

Some participants shared that when practicing at home with practice versions of necessary instruments and materials, they were only able to practice psychomotor skills and that there was no holistic patient experience. They liked the VR system as it simulated the actual ward with instruments and materials readily available. The high-resolution visuals were stimulating for the students as they performed the task. In addition, the VR procedures were rolled out during the COVID-19 pandemic, when in-person practice sessions were largely reduced. The lack of hands-on practice experience might be an impetus for students to embrace using technology for learning more readily. The students were keen to experience a new approach to learning and viewed it as a valuable learning opportunity.

In this study, most participants agreed that VR effectively allowed them to practice steps, enabled repeatable learning, and helped them form good habits, including hand hygiene habits, such as when they touched a patient or the environment with their hands and prior to performing any procedures. They were able to remain focused without being distracted by the external environment. In addition, IVR allowed the participants to proceed at their own pace and enabled them to construct new knowledge through the reflective thinking process. With the checklist embedded in the VR procedures, the students could receive timely feedback during their practice. Moreover, they had a sense of achievement when they completed the procedures successfully. The real-time feedback motivated the participants, and they strived to do better in each practice session. Such a feature is critical to enhancing learning, as Generation Z students look for instant results [4,5]. Most participants agreed that VR provided a good learning platform and that it helped them to be more focused on finer details to improve their competencies. The interactive clinical procedures were valuable as they encouraged self-regulated learning. The learners were able to develop new knowledge linked to previous learning when they eventually progressed from virtual practice to actual practice. This explains the willingness of undergraduate students to explore the IVR system, as it provided opportunities for them to put theory into practice and could improve their confidence in real-life practice [34].
Comparison to Prior Work
These findings are aligned with other studies showing that IVR is promising for improving engagement and motivating students to learn. Hence, this learning platform is suitable to be included as a teaching approach [36,37]. However, in this study, the ease of use and system quality of the VR procedures scored lower than the rest of the perception subscales. Similarly, the SUS also showed that more than half of the participants felt they required assistance to use the VR system, and that it was not easy to use. This signifies the importance of improving the efficiency and quality of the VR prototypes. The participants shared in their qualitative feedback that the controls were not intuitive and were confusing, as they were unsure which button to click. In addition, they reported technical issues, such as that the items kept sticking to their hands or dropping. This feedback helped the IT staff improve the procedures before implementing them for the entire cohort of students.

The participants expressed that they needed time to get used to and learn to use the VR device. Some used the VR set for a prolonged period, resulting in physical discomfort and cybersickness. This study highlighted that not all users find IVR comfortable, and that more refinement needs to be made to reduce technical difficulties. Therefore, IVR might not be suitable for everyone [38]. To ensure cyber safety, more precise tutorials and instructions should be developed to guide users. The user guide for the VR technology and the controllers will be revised. Maximum play duration and cautions and instructions for safety must be included in the orientation for users.

Finally, the participants also found the application lacking in audio and haptic feedback. They suggested including more cues by having vibration or buzzing when errors are made or having a visual dialogue box or audio prompt to inform players to increase interactivity. They shared that students would learn more if there were more challenging situations, such as including several types of intravenous solutions or insulins to stimulate students’ thinking. These ideas were valuable feedback for enhancing the learning experience, and the team has considered refining the prototypes.

Strengths and Limitations
The strengths of this study included, first, that all interactive clinical scenarios were reviewed and refined by the various parties. Second, faculty and students conducted several rounds of user acceptability testing to refine the prototypes prior to this study. Finally, developing the clinical procedures was a collaborative effort among clinical, academic, and IT professionals. All comments received were valuable, and improvements suggested by the participants were incorporated into the future refinement of the IVR procedures. The limitations included the IT professionals using Unreal Engine software for the first time, which led to teething issues. As they gained expertise in using the program, the IVR procedures significantly improved. Although the sample size was small, considering that it was a usability study, 29 participants was considered a sufficient size, and data collection for reviewing the design and evaluating its usefulness reached saturation. Going forward, more participants will be recruited to test the effectiveness of the IVR procedures.

Conclusions
IVR for simulating intravenous therapy and subcutaneous injection procedures was designed and developed for nursing students to perform the procedures at their own convenience and the time and place of their choosing. Our evaluation supports the feasibility of using IVR in learning clinical procedures. It highlighted the value of deliberate practice and scaffold learning via VR procedures. The study also supported the usability of the VR device, although improvements and revisions are needed for a better user experience. With the enhanced features, future research can examine the effectiveness of health care students’ learning clinical procedures at home using IVR. Given its portability and flexibility, IVR is potentially an effective adjunct to support health care students in learning and performing clinical procedures.

Acknowledgments
The authors would like to thank the collaborators and student participants who have supported this study. This study was supported by the Ministry of Education, Singapore, Tertiary Education Research Fund 2020 (MOE2020-TRF-053).

Data Availability
The data sets generated and analyzed during this study are not publicly available for privacy and ethical reasons but are available from the corresponding author upon reasonable request.

Authors’ Contributions
STL, SYL, JGY, and EA conceptualized and designed the study. STL, KDBR, FPL, and WLL conducted the data collection. STL, KDBR, and WLL carried out the data management, analysis, and synthesis of data. STL, RCJS, KDBR, and WLL wrote the study. All authors critically reviewed, read, and approved the final version of the study.

Conflicts of Interest
None declared.
References


2. Guze PA. Using technology to meet the challenges of medical education. Trans Am Clin Climatol Assoc 2015;126:260-270 [FREE Full text] [Medline: 2630687]


**Abbreviations**

- **IVR**: immersive virtual reality
- **SUS**: system usability scale
- **VR**: virtual reality

©Siew Tiang Lau, Rosalind Chiew Jiat Siah, Khairul Dzakirin Bin Rusli, Wen Liang Loh, John Yin Gwee Yap, Emily Ang, Fui Ping Lim, Sok Ying Liaw. Originally published in JMIR Serious Games (https://games.jmir.org), 30.08.2023. 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.
Effects of Exergaming on Musculoskeletal Pain in Older Adults: Systematic Review and Meta-analysis

Nan Mo1, MSc; Jin yu Feng1, MSc; Hai xia Liu1, MSc; Xiao yu Chen1, MSc; Hui Zhang1, MSc; Hui Zeng1, PhD
Xiangya Nursing School, Central South University, Changsha, China

Corresponding Author:
Hui Zeng, PhD
Xiangya Nursing School
Central South University
Xiangya School of Medicine
Number 172, Tong zipo Road
Changsha, 410013
China
Phone: 86 1 378 615 5688
Email: zenghui@csu.edu.cn

Abstract

Background: Exercise is effective for musculoskeletal pain. However, physical, social, and environmental factors make it difficult for older adults to persist in exercising. Exergaming is a new pathway that combines exercise with gameplay and may be helpful for older adults to overcome these difficulties and engage in regular exercise.

Objective: This systematic review aimed to determine the efficacy of exergaming to improve musculoskeletal pain in older adults.

Methods: The search was performed in 5 databases (PubMed, Embase, CINAHL, Web of Science, and Cochrane Library). The risk of bias for randomized controlled studies was assessed using the revised Cochrane Risk of Bias tool in randomized trials (RoB 2), and the methodological quality was assessed using the Physiotherapy Evidence-Based Database scale. Standardized mean difference and 95% CI were calculated using fixed-effects model meta-analyses in the Review Manager version 5.3 (RevMan 5.3).

Results: Seven randomized controlled studies were included, which contained 264 older adults. Three of the 7 studies reported significant improvements in pain after the exergaming intervention, but only 1 reported a significant difference between groups after adjustment for baseline (P<.05), and another reported a significant improvement in thermal pain between the 2 groups (P<.001). The results of the meta-analysis of the 7 studies showed no statistically significant improvement in pain compared to the control group (standardized mean difference –0.22; 95% CI –0.47 to 0.02; P=.07).

Conclusions: Although the effects of exergames on musculoskeletal pain in older adults are unknown, exergame training is generally safe, fun, and appealing to older adults. Unsupervised exercise at home is feasible and cost-effective. However, most of the current studies have used commercial exergames, and it is recommended that there should be more cooperation between industries in the future to develop professional rehabilitation exergames that are more suitable for older adults. The sample sizes of the studies included are small, the risk of bias is high, and the results should be interpreted with caution. Further randomized controlled studies with large sample sizes, high quality, and rigor are needed in the future.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42022342325; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=342325

KEYWORDS
aged; exergaming; pain; review; video game; virtual reality
Introduction

Background

The International Association for the Study of Pain defines pain as an unpleasant sensory and emotional experience associated with or resembling actual or potential tissue damage [1]. Pain is classified as acute, subacute, and chronic. Fewer than 4 weeks is acute pain, and the presence of 4 weeks to 3 months is subacute pain, and chronic pain is a condition that persists or recurs for more than 3 months [2]. The causes of pain in older individuals are often due to osteoarthritis, postherpetic neuralgia, diabetic neuropathy, spondylosis and radiculopathy, poststroke pain, and Parkinson disease [3]. According to data from the 2016 American Health Interview Survey [4], 20.4% of American adults have chronic pain, with 8% of American adults having a high impact on chronic pain [5]. The prevalence of pain increases with age [6]. Individuals aged 60 years and older are classified as older individuals by the World Health Organization [7]. Data from 1999 to 2019 showed that 57%-61% of community-dwelling older individuals reported intermittent or daily musculoskeletal pain [8]. By 2030, approximately 66% of people over the age of 65 years will have chronic pain globally [9]. Pain affects sleep and mood, increases the risk of falls, and reduces the quality of life [9-11]. It imposes a heavy burden on society [12].

Changes in the efficacy of analgesic drugs occur due to aging, such as a possible weakening in analgesic effect and a decrease in the efficiency of drugs acting on peripheral sensitization [13]. It is worth noting that medications may make older individuals more debilitated and adverse reactions occur more frequently [3]. Nonpharmacological treatments have therefore been used for pain relief. The efficacy of exercise in alleviating pain has been demonstrated [14], and exercising in nonpainful areas of the body has an analgesic effect on painful areas [15]. Older adults with poor physical function prefer to exercise at home at no cost [16]. However, the lack of supervision and motivation at home leads to low exercise adherence [17], which further leads to poorer treatment effects [18,19].

Technology can serve as an effective strategy to confront these challenges. Exergames are video games or virtual reality (VR) games that combine gameplay with physical training and are potential tools to make exercises more enjoyable and increase motivation and compliance for physical activity [20-23]. Results of systematic reviews showed that exergames could improve the activities of daily life [24], cognitive [25] and physical function [26], balance [27], walking speed [28], and depression [29,30] among older adults. There was some evidence of randomized controlled trials (RCTs) supporting the benefits of exergames for improving pain in older adults [31,32].

Research Gap and Aim

To our knowledge, 3 reviews have systematically summarized the effects of exergames on pain. A systematic review [33] included thirteen clinical studies, and the mean age of participants ranged from 23.9 (SD 6.8) years to 54.9 (SD 11.8) years. The 6 included controlled trials showed that interactive VR exergames may divert attention from pain and alleviate pain postmastectomy and ankylosing spondylitis, but the results were inconsistent for people with neck pain. The remaining 7 uncontrolled studies showed that interactive VR exergames reduced neuropathic limb pain and phantom limb pain, but did not affect nonspecific chronic back pain. A systematic review and meta-analysis [34] also showed that exergames can improve pain perception in females older than 18 years with fibromyalgia. However, a systematic review and meta-analysis [35] that included 7 RCTs concluded that there was insufficient evidence that exergames can improve musculoskeletal pain in the participants and the mean age ranged from 33.5 (SD 9.5) years to 80 years. The results of these reviews are inconsistent, as well as have some limitations. In the first place, they were not focused on older adults. Secondly, the included studies were not all RCTs. In further, some of them did not perform a meta-analysis and the results were not rigorous enough. Therefore, the purpose of this study is to review the efficacy of exergames for musculoskeletal pain in older adults.

Methods

Overview and Registration

The report of this systematic review and meta-analysis is consistent with the updated guidelines of the PRISMA (Preferred Reporting Items of Systematic Reviews and Meta-Analyses) 2020 Statement (Multimedia Appendix 1) [36]. The registration number is CRD4202342325.

Literature Search

A systematic literature search was carried out in 5 databases, PubMed, CINAHL (through EBSCO), the Cochrane Library, Web of Science, and Embase, from the inception to March 4, 2022. The combinations of Medical Subject Headings and free-text terms were used, and concepts included were exergaming, pain, and aged (see Multimedia Appendix 2).

Eligibility Criteria

The following were the criteria for including the articles: (1) participants’ mean age was more than 60 years and they suffered from musculoskeletal pain; (2) game technology was used to enable participants to exercise; (3) the control group was either active control (other interventions but no gameplay) or passive control (eg, usual care, no treatment, or waiting list); (4) the pain was involved in clinical outcomes; (5) the article had been published in a peer-reviewed publication with a RCT; and (6) the articles were written in English.

Exclusion criteria were as follows: (1) reviews, editorials, conference abstracts, and protocols, or full text was not available; (2) incomplete information on the intervention; (3) outcome data for pain were not statistically analyzed; and (4) duplicate publications or no restrictions on the publication date.

Study Selection

A researcher searched the 5 databases according to the search strategies. Duplicates were excluded by EndNote 9X. Two researchers independently reviewed the titles and abstracts of records before reading the entire text for rescreening to identify the included literature based on the eligibility criteria. Any differences were settled through discussion or by consulting a third researcher.
Data Extraction
The data were extracted by 2 independent researchers using the self-developed form in Excel (Microsoft Corp), comprising study characteristics, participant characteristics, intervention details, attrition, supervision, adverse events, experience, measurement tools for the outcome, and the key results. Following that, the 2 researchers subsequently cross-checked. A third investigator was consulted in the event of a dispute.

Risk of Bias and Methodological Quality Assessment
The included studies’ quality was evaluated using the revised Cochrane Risk of Bias tool in randomized trials (RoB 2) [37]. The risk of bias in 5 domains, including (1) randomization procedure; (2) deviations from intended interventions; (3) missing outcome data; (4) measurement of the outcome; and (5) selection of the reported result, was appraised using 3 degrees of “low risk,” “some concerns,” or “high risk.” When at least one domain was considered to have “some concerns,” but no domain was deemed to have “high risk,” the study was labeled as having “some concerns of bias.” When at least one domain was deemed “high risk” and many domains were deemed “some concerns,” the study was deemed “high risk of bias” [37].

The Physiotherapy Evidence-Based Database scale was used to assess the methodological quality of randomized controlled studies. The scale is a specific instrument for clinical studies of physical therapy interventions [38]. It comprises 11 items related to the selection, detection, performance, information, and attribution bases domains. Research with a score lower than 4 is regarded as bad, 4-5 is considered fair, 6-8 is considered good, and 9-10 is considered excellent [39].

The assessments were conducted independently by 2 independent evaluators. One reviewer resolved the disagreement.

Data Synthesis
The equation \(\text{Mean}_{\text{change}} = \text{Mean}_{\text{after}} - \text{Mean}_{\text{baseline}}\) and \(\text{SD}_{\text{change}} = \sqrt{\text{SD}^2_{\text{baseline}} + \text{SD}^2_{\text{after}} - 2 \times \text{correlation} \times \text{SD}_{\text{baseline}} \times \text{SD}_{\text{after}}}\) were used to calculate the mean change and corresponding SD, and the correlation was set to 0.5. SDs were not given in the study and were obtained by converting the means, sample sizes, and \(P\) values of the changes in the intervention and control groups [40]. The effect size was measured by the standardized mean difference (SMD) corrected for small sample sizes (Hedges’ \(g\)) [40]. Hedges’ \(g\) estimates of \(<0.30\), \(\geq0.30\) and \(<0.60\), and \(\geq0.60\) were considered small, moderate, and large, respectively [41]. The heterogeneity among studies was quantified based on the \(I^2\) statistic, with 0%-40% may not be important; 30%-60% may represent moderate heterogeneity; 50%-70% may represent substantial heterogeneity; and 75%-100% may represent considerable heterogeneity [42]. A fixed-effects model was used when \(I^2 \leq 50\%\), otherwise a random-effects model was used. We performed subgroup analysis to explore which treatments were more effective or what nature of pain exergaming was more effective for. Meta-analysis was performed using Review Manager version 5.3 (RevMan 5.3). We did not perform a publication bias test because fewer than 10 studies were included [42].

Results
Study Selection
By searching the databases, 2368 records were found. After removing 874 duplicates, 1494 records were evaluated. A total of 1411 records were excluded according to the eligibility criteria. The remaining 7 studies were analyzed. The PRISMA flow diagram depicts the search and screening procedure (Figure 1).
Quality Assessment

Figures 2 and 3 summarized the risk of bias assessment for the 7 studies that were included. One study (14%) was classified as “low risk” [44], 2 studies (29%) as “some concerns” [31,32], and 4 studies (57%) as “high risk” [21,22,45,46]. All studies were judged “low risk” for the domains “Deviations from intended interventions,” “Missing outcome data,” and “Selection of the reported result.” Four studies (57%) were assessed as “high risk” for the domain “Randomization procedure” due to no report allocation concealment approaches, while 1 research (14%) was evaluated as “some concerns” related to baseline imbalance. Due to nonblind assessors, 2 studies (29%) were determined to have “some concerns” in the domain “Measurement of the result.”
Figure 2. Risk of bias based on revised Cochrane Risk of Bias tool in randomized trials (RoB 2) [37].

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Randomization process</th>
<th>Deviations from intended interventions</th>
<th>Missing outcome data</th>
<th>Measurement of the outcome</th>
<th>Selection of the reported result</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beltran-Alacreu 2022</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dichburn 2020</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Fung 2012</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hsu 2011</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Monteiro-Junior 2015</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stamm 2022</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Zadro 2019</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 3. The overall risk of bias in randomized controlled trials.

Physiotherapy Evidence-Based Database scores (mean score approximately 7, range 6-8) demonstrated good overall methodological quality. Monteiro-Junior et al [44] received the highest score of 8, while Stamm et al [46] had the lowest score of 6 (Table 1).
Table 1. Methodological quality as assessed by the Physiotherapy Evidence-Based Database Scale. Item 1 did not count toward the total score.

<table>
<thead>
<tr>
<th>Study, year</th>
<th>1a</th>
<th>2b</th>
<th>3c</th>
<th>4d</th>
<th>5e</th>
<th>6f</th>
<th>7g</th>
<th>8h</th>
<th>9i</th>
<th>10j</th>
<th>11k</th>
<th>Total score</th>
<th>Methodological quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beltran-Alacreu et al [22], 2022</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Ditchburn et al [31], 2020</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Fung et al [45], 2012</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Hsu et al [21], 2011</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Monteiro-Junior et al [44], 2015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Stamm et al [46], 2022</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Zadro et al [32], 2019</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>

a1: Eligibility criteria.
b2: random assignment.
c3: allocation concealment.
d4: baseline comparability.
e5: subject blinding.
f6: therapists’ blinding.
g7: assessor blinding.
h8: adequate follow-up.
i9: intention-to-treat analysis.
j10: comparisons between groups.
k11: point estimates and variability.

Characteristics of the Included Studies

Study Characteristics

Seven included studies were published in years from 2011 to 2022. They were conducted in Spain [22], England [31], Canada [21,45], Brazil [44], Germany [46], and Australia [32]. Study designs were RCTs (n=4) [31,32,44,45], pilot RCTs (n=1) [46], and crossover pilot RCTs (n=2) [21,22]. The sample size ranged from 14 to 60, with a total of 264, of which 137 were in the intervention groups and 127 were in the control groups.

Participant Characteristics

All participants had musculoskeletal pain, participants in 5 studies [22,31,32,44,46] had chronic pain, and participants in the 2 studies [21,45] had nonchronic pain. The average age of the participants ranged from 67.8 (SD 6) years [32] to 81.85 (SD 6.82) years [22]. The majority of participants were female, accounting for around 69% of the total.

Intervention Characteristics

Interventions were conducted in the nursing home [22], the university’s physiotherapy laboratory [31], the hospital [45], the long-term care center [21], the center of rehabilitation [44], the laboratory [46], and the participant’s home [32]. The Nintendo Wii, the Active Airlines serious game, the Interactive Rehabilitation and Exercise System, and the ViRST VR application were the main gaming platforms in the experimental groups. For a period of 4 weeks to 8 weeks, participants exercised for 210 seconds to 90 minutes every session, only once during the intervention to 3 times per week. In 1 study, participants in the control groups carried on with their regular activities. The other studies used traditional physical therapy.

In 4 studies, participants performed exergaming under the supervision of the first author [31], therapist [21,45], and physiotherapist [46]. In 1 study [32], participants performed unsupervised home exergaming, and the remaining 2 studies did not report whether supervision was implemented [22,44]. The number of attrition people ranged from 1 to 9 due to personal commitments and fear of COVID-19 infection. Adherence was reported in only 3 studies [21,32,44].

Most of the participants in the experimental group had a positive experience. The occurrence of adverse events was not reported in 3 studies [31,44,45], while 3 studies reported no adverse events during the intervention [21,32,46]. Two individuals in 1 study experienced unpleasant symptoms such as dizziness, eye pain, or disorientation [22]. Multimedia Appendix 3 depicts an overview of the included study characteristics.

Results of Studies

The Effect of Exergames on Pain

The complete pain data was presented in 7 papers. A study [22] comparing the effects of using Active Airlines serious game and conventional exercise on pain using Visual Analog Scale (VAS) measurements after 4 weeks discovered that both experimental and control groups had significant improvements in chronic neck pain, but exergaming therapy was not superior to conventional exercise. Ditchburn et al [31] compared exercise using the VR rehabilitation system to traditional gym-based exercise, and the results of the study, measured at baseline and after 6 weeks using the VAS, showed a significant improvement in chronic musculoskeletal pain intensity in the experimental group, but no significant change in the control group, and no statistically significant difference between the 2 groups. The difference in improvement in thermal pain, including burning and hot, measured by Multi Affect and Pain Survey, was
significant between the 2 groups. In the study by Stamm et al [46], the results measured at baseline versus 6 weeks later on the Numeric Rating Scale (NRS) showed that the improvement in chronic back pain was not significant in the experimental group using the VR system and those in the control group receiving traditional pain therapy, and the difference between the 2 groups was not statistically significant. Fung et al [45] investigated the effect of Nintendo Wii Fit gaming sessions on pain in posttotal knee replacement individuals, finding no significant difference between the 2 groups when compared to a control group receiving lower extremity exercise. Two research investigations the effects of exergaming combined with traditional exercise with regular exercise on pain perception, with Nintendo Wii used. Hsu et al [21] used the NRS and the pain bothersomeness of the upper extremity to measure the improvement of pain in people with upper extremity dysfunction in the experimental and control groups after 4 weeks of treatment and found no significant differences. Another study [44] reported significant improvement in chronic low back pain after 8 weeks of treatment in participants in both the experimental and control groups using the NRS, however, the difference between the 2 groups was not statistically significant. The study by Zadro et al [32] evaluated the effect of home-based exercise via Wii Fit U on chronic low back pain, measured by NRS at baseline versus 8 weeks later, with no statistically significant difference compared to usual activities, however, after adjustment for baseline, the results showed a statistically significant difference between the 2 groups.

A meta-analysis of data provided by VAS and NRS in 7 studies was performed and the overall results of this meta-analysis were inconclusive (SMD = -0.22, 95% CI = -0.47 to 0.02; P = 0.07; I² = 0%; the fixed-effect model; Figure 4).

**Figure 4.** The effects of exergaming on pain perception.

<table>
<thead>
<tr>
<th>Study</th>
<th>Exergame group</th>
<th>Control group</th>
<th>Std. Mean Difference</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>125.0</td>
<td>17.1</td>
<td>117</td>
<td>125.0</td>
</tr>
<tr>
<td></td>
<td>-0.22</td>
<td>-0.47</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

**Subgroup Analysis**

We investigated the differences in the effects of pain perception on the comparisons. Subgroup analysis showed no significant effect of exergaming in combination with traditional physical therapy (SMD = -0.04; 95% CI = -0.53 to 0.45; P = 0.87; I² = 0%; the fixed-effects model; Figure 5) or exergaming alone (SMD = -0.2; 95% CI = -0.54 to 0.13; P = 0.24; I² = 0%; the fixed-effects model; Figure 5) on pain perception compared to traditional physical therapy. There was no statistically significant difference between exergaming compared to usual activities (SMD = -0.47; 95% CI = -0.98 to 0.04; P = 0.07; I² = 0%; the fixed-effects model; Figure 5). The effect sizes for the 3 comparisons were small (<0.3), medium (≥0.3 and <0.6), respectively.

**Figure 5.** The effect on pain perception in the different comparisons.

<table>
<thead>
<tr>
<th>Study</th>
<th>Exergame group</th>
<th>Control group</th>
<th>Std. Mean Difference</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>125.0</td>
<td>17.1</td>
<td>117</td>
<td>125.0</td>
</tr>
<tr>
<td></td>
<td>-0.22</td>
<td>-0.47</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

https://games.jmir.org/2023/1/e42944
We conducted subgroup analyses to investigate the effects of exergaming on pain perception in participants with chronic pain or nonchronic pain. Subgroup analyses of the 7 studies revealed a nonsignificant difference in the effect of exergaming on chronic pain compared to controls (SMD -0.16; 95% CI -0.45 to 0.14; $P=0.29$; $I^2=0$%; the fixed-effects model; Figure 6). Similarly, there was no statistically significant effect of exergaming on nonchronic pain compared to the control group (SMD -0.36; 95% CI -0.80 to 0.07; $P=0.01$; $I^2=0$%; the fixed-effects model; Figure 6). Nevertheless, the effect sizes were small for chronic pain (<0.3) and moderate for nonchronic pain (≥0.3 and <0.6). We proceeded to investigate in depth the effect of exercise frequency on pain perception in participants with chronic pain. Subgroup analysis of the 5 studies showed no significant difference in the effect of exercise frequency of twice a week (SMD -0.16; 95% CI -0.64 to 0.32; $P=0.51$; $I^2=0$%; the fixed-effects model; Figure 7) and 3 times a week (SMD -0.16; 95% CI -0.53 to 0.22; $P=0.41$; $I^2=35$%; the fixed-effects model; Figure 7) on chronic pain compared to the control group, and the effect size was the same for both exercise frequencies (0.16).

Figure 6. The effects of exergaming on chronic pain or nonchronic pain.

Figure 7. The effect of exercise frequency on chronic pain.

Discussion

Efficacy

This meta-analysis and systematic review focused on the effects of exergaming on musculoskeletal pain in older adults and included 7 randomized controlled studies. The main finding is that the effect of exergaming on musculoskeletal pain in older adults is inconclusive. The results of this study are similar to Collado-Mateo et al [35], and they concluded that exergaming is more difficult to improve musculoskeletal pain in older individuals compared to adults. Of the 7 studies we included, 1 study [32] adjusted for baseline reported a significant improvement between groups. Another study [31] showed significant differences between the experimental and control groups in terms of improvement in thermal pain. Furthermore, the quality of the evidence is low and the sample sizes in the studies were quite small. Further research is needed on the effect of exergames on musculoskeletal pain in older adults. The highest mean effect size of -0.47 was observed for improvement in pain in the comparison of exergaming versus
usual activities, although it was not statistically significant. In this comparison, only 1 study [32] was included and adjusted for baseline, which showed statistical significance. This suggests that exergames have some potential in the treatment of musculoskeletal pain in older adults and that future high-quality studies are needed. Exergaming combined with traditional physiotherapy compared to traditional physiotherapy alone yielded the lowest mean effect size of –0.04, but still favored the experimental group despite the nonsignificant difference between the groups. The results of the 2 included studies were contradictory. Hsu et al [21] examined older adults with upper limb dysfunction, and after 4 weeks, the pain did not improve within either group. In contrast, in the study by Monteiro-Junior et al [44], the pain was significantly improved within both groups after 8 weeks, but the difference between the groups was not significant and showed results in favor of the control group. First, the difference in the results of the 2 studies may be related to the disease experienced by the participants, with different duration of the intervention. Second, the results of the study by Monteiro-Junior et al [44] favored the control group, possibly because participants in the experimental group had to complete not only traditional strength and core training, but also exergame training through the Nintendo Wii, with a total training time of 90 minutes each time, 3 times per week, and higher exercise intensity, which may have made people difficult to obtain optimal results. The results of the network meta-analysis by Fernández-Rodríguez et al [47] suggested that core exercises, strength exercises, or mind-body exercises for less than 60 minutes at a time, at least once to twice a week, with exercise lasting 3 to 9 weeks, are the most beneficial treatment for pain and disability in adults with chronic lower back pain exercise program. Due to older adults tend to have lower endurance levels, they are more susceptible to sports injuries and overexertion and have difficulty tolerating high-intensity training. The sample sizes of the 2 studies were small and the results should be interpreted with caution. In the comparison of exergaming with traditional physical therapy, the mean effect size was –0.2. The results favored the experimental group. Overall, exergaming can be used as adjunctive alternative therapy to traditional physical therapy.

These studies included older adults with a variety of musculoskeletal pain. Most of the participants in the studies suffered from chronic pain (k=5) such as back pain (k=3), neck pain (k=1), musculoskeletal pain (k=1), and other nonchronic pain such as upper extremity dysfunction (k=1) and post total knee replacement (k=1). These can be explained by the results of epidemiological studies, in which the most common pain complaints were osteoarthritis back pain, especially in the low back or neck (65%), musculoskeletal pain (40%), peripheral neuropathic pain (35%), and chronic joint pain (15%-25%) [9]. The results from the subgroup analysis showed that the effect sizes of exergaming on improving nonchronic pain were greater than the effect sizes on improving chronic pain. The results of this study are inconsistent with those of Collado-Mateo et al [35] and the results may be due to age-related group differences, and our study focused only on the group of older individuals. More studies are needed in the future.

The mechanisms by which exercise ameliorates pain are unclear, with 1 suggestion being that exercise leads to an increase in stress pain thresholds and that adaptation of central inhibition occurs over time with exercise training [48]. A meta-analysis showed that increasing the frequency of weekly exercise was most likely to have a positive impact on patients with chronic pain [49]. However, our study results showed that exercise frequency of twice a week and 3 times a week had the same size of effect on chronic pain. There are no standardized criteria for exergaming intervention programs, and it is particularly important to develop an appropriate exercise program. From the RCTs included in this study, 4 weeks of the exercise was sufficient to significantly improve pain, at least twice a week, but not for more than 90 minutes per session.

Exergames Design

Most of the 7 studies tested commercial game platforms, with 1 study using a training rehabilitation-specific platform [31], and participants in the exergaming group experienced significant improvements in pain and for thermal pain, there was a significant difference between the 2 groups, the only 1 of the included studies to show a significant between-group difference for improvement in pain. Therefore, using professional rehabilitation exergames may be more effective than commercial games [23]. Professional exercise rehabilitation games are more specialized because they may be developed with the involvement of professionals in their design and can take into account the type of illness the users have, their needs, etc. Most commercial exergames are not suitable for the group of older individuals, about speed, required movements, amount of information, etc [50]. Therefore, in the future, commercial and medical rehabilitation professions should strengthen their cooperation to develop user-centered exergames for older individuals [51], thus improving the efficacy of exercise [52]. For older individuals, their physiological characteristics [53] and motivations for use should be considered. Older adults are motivated more by perceived health effects, the pleasure of the game, and the improvement of social confidence [20,54]. Wang et al [55] suggested that when designing exergames, first, aging characteristics should be included, paying attention to the decline of cognitive and physiological abilities associated with aging. Second, the game motion recognition should have higher fault tolerance. Third, the feedback should be clear. Fourth, consider the endurance of older individuals, pay attention to fatigue management and control the pace of the game. Fifth, it should have continuous action cues and tutorials. Sixth, it should be connected with reality. Seventh, reasonable use of body parts. Eighth, make good use of repetitive actions and reversible actions. Ninth, design advanced actions for the same game tasks. Lastly, designers should take advice from rehabilitation experts when designing exergames.

Supervision and Adherence

Except for Zadro et al [32], who studied unsupervised home exergame training and found a significant effect of exergaming on pain after adjusting for baseline, the majority of research participants were supervised during exergaming. Compliance among participants was higher but still lower than with supervised exergame training. A prior study indicated that while
home exercise training relieved low back pain, supervised training improved pain intensity the greatest [56]. However, to the best of our knowledge, no studies have been conducted to compare unsupervised home exergaming with unsupervised home exercise in pain relief for older individuals. Exergaming, in general, remains a highly promising kind of training that allows participants to undertake unsupervised therapeutic exercises at home, capable of generating a remote rehabilitation environment. Older individuals who are frail or incapacitated can obtain therapy without having to travel vast distances, which may have significant cost-effectiveness benefits [22]. Because just 1 study on home exercise was included in this paper, the results were insufficiently persuasive. As a result, further research might be done in the future, and methods to promote adherence to unsupervised exergame training in older people at home could be pursued.

Security and Experience
Exergames are generally safe for older adults, although a few participants reported feeling uncomfortable, which may be related to the device and form of movement, such as wearing sensors to move the neck making participants feel dizzy, uncomfortable with their eyes, and disoriented [22]. From the reports of coaches and participants, it was found that exergames increased the fun and attraction of physical activity and made the game more enjoyable for the participants [21,45,46]. Previous research has also concluded that participants in the exergaming group had significantly more enjoyment of exercise than the other treatment groups [57]. Participants found exergames highly usable and the game challenging [22]. Exergames can increase participants’ satisfaction and compliance [22]. Some studies [20,46] considered socialization as an important factor in improving adherence, as stated in previous studies [58]. However, participants in a control group in 1 study showed a higher acceptance of traditional exercise than exergames [31], possibly because participants in the control group had not experienced exergames. Overall, older adults have positive attitudes toward exergaming.

Limitations
It is the first systematic review of the efficacy of exergaming on musculoskeletal pain in older adults. Several limitations should be in consideration. When searching the literature, publications were limited to those in English and peer-reviewed ones. The small sample sizes of most studies were also a limitation. In addition, the overall risk of bias in the included studies was relatively high.

Conclusions
This paper systematically reviews the efficacy of exergaming on musculoskeletal pain in older adults. The available evidence is limited, and therefore, exergaming cannot yet be considered an effective intervention for improving pain in older adults. Exergames are safe and cost-effective. The playfulness and social components of exergaming may contribute to participant adherence. Increased collaboration between industries to develop specialized exergames for older adults should be considered in the future. Overall, exergaming can be used as a complementary alternative to traditional training. Future larger sample sizes and rigorously designed RCTs are needed to explore the effects of different exergames on older adults with musculoskeletal pain.

Acknowledgments
This study was funded by the Hunan Provincial Health Commission (202114052087), and the Department of Science and Technology of Hunan Province (2021ZK4163).

Authors’ Contributions
NM contributed in conceptualization, literature search, quality evaluation, data extraction, and writing original draft. JF was involved in literature search, quality evaluation, data extraction, and writing, review, and editing. HL was responsible for literature search, quality evaluation, data extraction, and writing, review, and editing. XC performed literature search, quality evaluation, data extraction, and writing, review, and editing. H Zhang was responsible for writing, review, and editing. H Zeng contributed in writing, review, editing; supervision; and project administration.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[DOCX File, 60 KB - games_v11i1e42944_app1.docx ]

Multimedia Appendix 2
Search strategy.
[DOCX File, 33 KB - games_v11i1e42944_app2.docx ]

Multimedia Appendix 3
Characteristics of included studies.

https://games.jmir.org/2023/1/e42944
References


50. Brox E, Konstantinidis ST, Evertsen G. User-Centered Design of Serious Games for Older Adults Following 3 Years of Experience With Exergames for Seniors: A Study Design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]

**Abbreviations**

NRS: Numeric Rating Scale

PRISMA: Preferred Reporting Items of Systematic Reviews and Meta-Analyses

RCT: randomized controlled trial

RoB 2: Revised Cochrane Risk of Bias tool in randomized trials

SMD: standardized mean difference
**VAS:** Visual Analog Scale

**VR:** virtual reality

---

Edited by G Eysenbach, N Zary; submitted 25.09.22; peer-reviewed by M Behzadifar, Y Wang, O Stamm; comments to author 28.12.22; revised version received 16.02.23; accepted 29.03.23; published 25.04.23.

**Please cite as:**
Mo N, Feng JY, Liu HX, Chen XY, Zhang H, Zeng H
Effects of Exergaming on Musculoskeletal Pain in Older Adults: Systematic Review and Meta-analysis
JMIR Serious Games 2023;11:e42944
URL: https://games.jmir.org/2023/1/e42944
doi:10.2196/42944
PMID:37097717

©Nan Mo, Jin yu Feng, Hai xia Liu, Xiao yu Chen, Hui Zhang, Hui Zeng. Originally published in JMIR Serious Games (https://games.jmir.org), 25.04.2023. 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.
Viewpoint

Time to Think “Meta”: A Critical Viewpoint on the Risks and Benefits of Virtual Worlds for Mental Health

Vincent Paquin1,2, MD; Manuela Ferrari1,3, MHSc, PhD; Harmehr Sekhon1,2,4, MSc, PhD; Soham Rej1,2,3, MD, MSc

1Department of Psychiatry, McGill University, Montreal, QC, Canada
2Lady Davis Research Institute, Jewish General Hospital, Montreal, QC, Canada
3Douglas Mental Health University Institute, Montreal, QC, Canada
4McLean Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Vincent Paquin, MD
Department of Psychiatry
McGill University
1033 Avenue des Pins
Montreal, QC, H3A 1A1
Canada
Phone: 1 514 398 4909
Email: vincent.paquin2@mail.mcgill.ca

Abstract

The metaverse is gaining traction in the general population and has become a priority of the technological industry. Defined as persistent virtual worlds that exist in virtual or augmented reality, the metaverse proposes to afford a range of activities of daily life, from socializing and relaxing to gaming, shopping, and working. Because of its scope, its projected popularity, and its immersivity, the metaverse may pose unique opportunities and risks for mental health. In this viewpoint article, we integrate existing evidence on the mental health impacts of video games, social media, and virtual reality to anticipate how the metaverse could influence mental health. We outline 2 categories of mechanisms related to mental health: experiences or behaviors afforded by the metaverse and experiences or behaviors displaced by it. The metaverse may benefit mental health by affording control (over an avatar and its virtual environment), cognitive activation, physical activity, social connections, and a sense of autonomy and competence. However, repetitive rewarding experiences may lead to addiction-like behaviors, and high engagement in virtual worlds may facilitate and perpetuate the avoidance of challenges in the offline environment. Further, time spent in virtual worlds may displace (reduce) other determinants of mental health, such as sleep rhythms and offline social capital. Importantly, individuals will differ in their uses of and psychological responses to the metaverse, resulting in heterogeneous impacts on their mental health. Their technological motivations, developmental stage, sociodemographic context, and prior mental health problems are some of the factors that may modify and frame the positive and negative effects of the metaverse on their mental health. In conclusion, as the metaverse is being scaffolded by the industry and by its users, there is a window of opportunity for researchers, clinicians, and people with lived experience to coproduce knowledge on its possible impacts on mental health and illness, with the hope of influencing policy-making, technological development, and counseling of patients.

(JMIR Serious Games 2023;11:e43388) doi:10.2196/43388

KEYWORDS

metaverse; digital media; virtual reality; mental health; addiction; social functioning; virtual; technology; augmented reality; gaming; social media; cognitive; physical activity; behavior; psychological; development; patient; policy

Introduction

In 2021, Facebook spent US $10 billion on its metaverse division, shipped 10 million virtual reality headsets, and rebranded itself as Meta [1]. The same year, 46 million users were logging onto the game platform Roblox every day, spending a total of 41 billion hours in this virtual world [2]. Concurrently, technology giants Apple, Microsoft, and Google have been developing their own metaverse-related infrastructures. While the idea of a metaverse is not new, the popularity of virtual worlds has boomed in the past few years, particularly during the COVID-19 pandemic [2,3].

The technology magazine WIRED defines the metaverse as an extension of “virtual reality—characterized by persistent virtual...
worlds that continue to exist even when you're not playing—as well as augmented reality that combines aspects of the digital and physical world” [4]. In virtual reality, users wear headsets to see a digital world in 3 dimensions. Through the avatars they embody, they can move in the virtual environment, have life-like interactions with other users, and control digital objects. In augmented reality, users see their real, physical environment, but digital elements are integrated through eyeglasses or a smartphone. Pokémon Go is an example of augmented reality: in this mobile game, the real-world locations of players are captured by their smartphone camera and overlaid with virtual creatures that they can interact with. Other devices combine elements of virtual and augmented reality, allowing users to simultaneously interact with virtual and real objects in “mixed reality” [5].

Although virtual worlds such as Pokémon Go and Roblox are already ubiquitous, the promise of the metaverse is to make these spaces more immersive and interconnected, and to extend their scope of application beyond play unto other aspects of daily life. Of late, there has been an emergence of virtual reality platforms such as VRChat, whose main purpose is to allow users to interact and socialize with others with their self-created 3D avatars. Increasingly, users of virtual worlds trade cryptocurrencies, which are digital currencies that function like money but are independent from states and central banks [6]. With cryptocurrencies, one can buy digital assets (eg, art, avatar apparel, and virtual estates) that are recorded on virtual ledgers as uniquely having one owner, and that are not interchangeable or duplicable; these assets are called nonfungible tokens [7]. Tying together a range of emerging technologies, metaverse platforms thus offer rich and vivid experiences, from games to new ways of socializing, working, learning, and shopping [4,8].

Will the metaverse impact mental health and illness? With every media revolution threatening to infiltrate people’s homes, whether the radio, TV, or the internet, there are concerns for mental health [9]. In line with popular views, a body of research points to potential harmful effects of digital media on depression, anxiety, and addiction-like behaviors [10-12]. However, in truth, cyberpsychology—the study of how humans and computers interact—paints a much more complex picture [13]. The mental health risks of digital media are far from universal and, at least in some cases, the adaptive use of technology may even contribute to improvement in short- and long-term well-being [14]. As interest in the possibilities of the metaverse continues to grow in the general population, we consider in the present viewpoint how this technology may impact mental health. We narratively present a selection of key articles from the literature on digital media to appraise the potential mental health risks and benefits of the metaverse and discuss directions for future research.

Virtual Worlds and Mental Health

Overview

Most research to date has examined exposure to digital media as a function of screen time, roughly defined as time spent on digital media and other screen devices. This body of literature generally shows that individuals who report greater screen time also report poorer mental health [15]. However, this association is likely to be confounded by other factors, such as preceding mental health conditions, occupational status, and lifestyle, to name a few. When individuals are studied as their own comparators, fluctuations in digital media use over time do not seem to predict changes in mental health [16-19], which goes against the hypothesis that higher screen time causes poorer mental health. However, the utility of screen time as an indicator of digital exposures is likely limited if the amount of time on digital media is not the primary mechanism by which digital media affect mental health. Media researchers have advocated for dropping the concept of screen time and focusing instead on more specific, theory-based measures of digital media exposures [14,16,20]. Limited research has investigated the mental health impacts of the metaverse, but many of the current platforms described as “metaverse” rely on virtual reality and are intended for internet-based gaming or socializing (see the Introduction section). The metaverse shares a number of characteristics with social media and video games: these include the possibility of interacting with other individuals in real time regardless of geographical location, the possibility of being anonymous, and the persistence of data over time [21,22]. Video games and virtual reality further afford vivid experiences that can be studied to better understand the potential psychological effects of immersion in the metaverse.

Thus, to anticipate how the metaverse will affect mental health, we can learn from the literature examining the mental health impacts of social media, video games, and virtual reality. From this state of knowledge, 2 broad categories of mechanisms, by which the metaverse may influence mental health, emerge: experiences or behaviors that are afforded by its use and experiences or behaviors that are displaced by it [9]. Below, we discuss these mechanisms in detail and summarize them in Figure 1.
Affordances of the Metaverse

Environmental Control and Self-representation

Virtual worlds of the metaverse have the potential to afford users a number of experiences susceptible of impacting their mental health. This begins with the environment of the person, on which the metaverse affords a unique degree of control: users can easily choose and navigate the environment of their choice, or even self-design the space they are in. Virtual reality relaxation exploits this affordance by bringing the user into restorative, pleasant environments such as lakes, forests, or rivers. According to a recent systematic review, there is preliminary experimental evidence that virtual reality relaxation is effective for improving mood and anxiety, at least in the short term [23]. Conversely, exposure to social stressors in virtual environments is associated with higher levels of subjective distress and paranoia [24], illustrating the potential influence of virtual environmental exposures on mental health.

Virtual worlds also afford users to choose their avatars’ appearance and identity, which can influence how they perceive and interact with their surroundings. In an international survey of 142 regular users of video games, greater embodiment over an avatar was associated with lower awareness of bodily sensation during gaming sessions [25]. In a sample of 60 women from the UK general population, reducing avatar height during a virtual reality train ride was associated with increased levels of paranoia and negative social comparison [26]. Experimenting with self-representation in the metaverse may also participate in youth development: some authors have suggested that during adolescence, conducting identity experiments on the internet could help in the development of social competence [27]. However, strong investment in self-representation in virtual spaces could also cause harms, such as dissatisfaction with one’s physical body. This risk has been documented in studies of social media usage, showing that people’s investment in receiving feedback on their self-portrait photographs, and their comparisons with the photographs of others, may contribute to body dissatisfaction and drive for thinness [28,29].

Cognitive Activation and Physical Activity

The design of metaverse contents may be harnessed toward activating a person’s cognitive functions in beneficial ways. Many games have been developed to encourage players to use their cognitive skills to solve problems [30]. In a randomized controlled trial of 72 US adults with major depressive disorder, a video game–based intervention was effective in improving sustained attention and global cognitive functioning compared with an active control [31]. The embodied navigation of virtual worlds, enabled by sensors that capture the user’s movement in physical space, can also promote physical activity. This is the case of video games involving physical activity.
Attributing problematic gaming behaviors as constitutive of a type being internet-based rather than offline gaming [43]. The prevalence of gaming disorder at 2%-3%, with the most common negative consequences. A meta-analysis situates the global and continuation or escalation of gaming despite the occurrence given to gaming (at the expense of other interests and activities), manifesting as impaired control over gaming, increased priority may translate into mental health problems [42]. Gaming disorder, which made its entry in the 11th revision of the International Classification of Diseases, captures how this 

Social Connectedness

Interpersonal relationships could be developed and maintained through the metaverse. The capacity to connect with others regardless of physical location enhance individuals' opportunities to join communities with shared interests, values, or identities. Members of marginalized groups and people living with mental health problems may benefit from these platforms for sharing knowledge and obtaining peer support [33-35]. Importantly, internet-based interactions can be experienced as meaningful. In an international survey of people using massively multiplayer online role-playing games (MMORPGs), respondents frequently reported high-quality relationships in their games, which sometimes transformed into friendships or romantic partnerships in the physical world [36,37]. MMORPG users also commonly play with their offline friends or romantic partners [37,38], suggesting that virtual worlds can foster both the formation of new relationships and the enhancement of existing ones. More research is needed to understand the extent to which these social affordances impact mental health, but arguably, they may be beneficial if they decrease a person’s loneliness and increase their social capital [39].

Self-determination

Experiences of autonomy and competence in virtual worlds can fulfill the need for self-determination. In a survey of 672 MMORPG players, positive experiences associated with intensive gaming involvement included feelings of achievement, positive anticipation and stress (eg, “adrenaline rush”), enjoyable repetition and routine, positive social obligation, satisfying labor, increased confidence, and positive distraction, to name a few [40]. In samples of college students playing video games in experimental settings, experiences of autonomy and competence while playing were associated with gaming motivation and enjoyment [41]. For people experiencing disability, functional impairments, or social adversity in the context of mental health problems, digital spaces that operate under different parameters than the offline, physical world may thus provide invaluable opportunities for fulfilling the psychological needs related to self-determination.

Addictions

Because they repetitively afford rewarding experiences, virtual worlds can also be the object of addiction-like behaviors. Gaming disorder, which made its entry in the 11th revision of the International Classification of Diseases, captures how this may translate into mental health problems [42]. Gaming disorder is defined by persistent or recurrent gaming behavior, manifesting as impaired control over gaming, increased priority given to gaming (at the expense of other interests and activities), and continuation or escalation of gaming despite the occurrence of negative consequences. A meta-analysis situates the global prevalence of gaming disorder at 2%-3%, with the most common type being internet-based rather than offline gaming [43].

Attributing problematic gaming behaviors as constitutive of a mental disorder is subject to controversy [10,44]. But at the minimum, epidemiological and cross-cultural data show that some players engage in gaming in ways that negatively impact their well-being and functioning [43,45]. In the metaverse, immersive technologies broaden the scope of daily-life activities that can be performed digitally, raising the possibility of even greater absorption into virtual worlds and greater interference with functioning. Addiction is a particular concern considering that companies have a financial incentive to maximize virtual worlds' capacity to captivate users—something that can be facilitated by passively monitoring users’ behaviors in the metaverse and by customizing their environment accordingly [46].

Displacement of Offline Activities

Circadian Rhythms

High engagement in the metaverse could be at the expense of other activities in the offline environment. According to the displacement hypothesis [47], excessive use of digital media causes harm to the individual by displacing time spent on other activities that are beneficial for mental health. Notably, participating in virtual worlds in the evening or at night may impair sleep rhythms because of delayed or reduced sleep time. According to a recent narrative review, previous observational studies support an association between greater use of screen-based media and poorer sleep outcomes in adolescents; however, experimental research does not consistently show that screen use has causal effects on sleep [48]. Three mechanisms could be responsible for the putative effects of screen use on sleep: disruption of the neurobiology of sleep due to light emitted by the screen devices, delay of sleep onset or reduction of total sleep time because of time spent on the screen device, and psychological stimulation caused by digital media content [48]. More evidence is needed to determine the causal effects of screens on sleep and whether the degree of immersion afforded by the metaverse will amplify these effects.

Offline Stress Exposure

Individuals may participate in virtual worlds to avoid or escape from anxiety-provoking situations of the offline world. In web-based surveys, avoidance and escapism are relatively commonly reported by video gamers as part of their motivations for gaming [36,45,49]. For example, in an international survey of internet-based gamers, participants endorsed some degree of “gaming to avoid challenges in [their] life rather than deal with them directly” (average item score of 2.46 on a scale of 1 to 5)—a finding that was consistent across North America, Europe, and China [45]. Recurrent avoidance of stressful situations may prevent a person from seeking to solve the problems or from learning to tolerate the stressors they face. However, it has been suggested that escapism through video gaming can also be an adaptive coping strategy that helps regulate or restore mood following exposure to stressful situations [50,51]. A systematic review of surveys conducted among video gamers found that escapism is cross-sectionally associated with both negative outcomes (eg, problematic video gaming, social anxiety, and loneliness) and positive outcomes (eg, social connection, enjoyment, and psychological well-being) [49]. Longitudinal studies are needed to better understand how escapism in the
metaverse interacts with and affects a person’s mental health problems over time.

**In-Person Interactions**

A related topic is how the potential displacement of in-person social interactions toward their digital equivalent influences mental health. As highlighted above, there is evidence that internet-based gaming can foster meaningful social interactions both digitally and offline. However, if high social involvement in the metaverse is accompanied by limited social interaction in the offline world, the net result may be an increase in social anxiety and a decrease in social skills in offline contexts [39]. This could explain why internet addiction was associated with a prospective increase in loneliness in a convenience sample of 361 college students in Hong Kong [52]. Potentially, for individuals living with social anxiety or low social skills, achieving connectedness through virtual worlds can simultaneously be a compensatory strategy and a perpetuating factor that constrains exposure to in-person social interactions.

**The Importance of Individual and Contextual Factors**

**Overview**

The effects of virtual worlds on mental health likely depend on individual and contextual factors. To begin with, virtual worlds and the technologies that support them differ in their features and affordances, including those discussed above: control of the avatar and environment, cognitive stimulation, social connectedness, and the satisfaction of other psychological needs. Users also have freedom in how they take advantage of virtual worlds’ possibilities. Differences in digital media uses, and their relevance for mental health correlates, are frequently observed in the scientific literature. For example, in an experience sampling study of 44 adults with and those without psychosis in the United Kingdom, posting about feelings and venting on social media predicted higher subsequent paranoia, while posting about daily activities predicted lower paranoia [53]. Although such findings are correlational and subject to confounding, they reflect the heterogeneity of uses and mental health correlates for a given technology. This heterogeneity is perhaps best understood in terms of individual or contextual factors. Borrowing on the terminology of Valkenburg and Peter [54], factors that modify and frame the mental health impacts of digital media can be organized in 3 categories: dispositional, developmental, and social.

**Dispositional and Developmental Factors**

At the dispositional level, individuals’ motivations and interests will likely influence how they engage with the metaverse. Among US school-aged children, boys generally report greater use of video games than girls, whereas girls report greater use of social media [55]. In an international longitudinal study of video game players, a person’s increase in intrinsic motivation for gaming was associated with a subsequent improvement in their affect and life satisfaction; conversely, an increase in extrinsic motivation (ie, feeling pressured to play) was association with a deterioration of the same outcomes [16]. At the developmental level, age may moderate the mental health outcomes of metaverse usage as a function of the person’s developmental needs and sensitivities, as well as age-dependent differences in motivations and profiles of use. To illustrate, early childhood is a crucial period for cognitive development, and there is tentative evidence from a Canadian study that a child’s increase in screen time is associated with a decrease in their performance on developmental screening tests between ages of 24 and 60 months [56]. Adolescence is marked by social and identity development, and teens may be particularly sensitive to social acceptance and rejection in virtual worlds [57]. Some authors have also argued that digital experiences have the potential to contribute to agency, connectedness, and storytelling capacities that foster youth’s identity development [58].

**Social Factors**

Social and economic factors are sources of inequalities in the access to metaverse-related technologies. Lower income or economic development is associated with reduced access to internet-connected devices, and in many emerging countries, women have lower access to the internet than men [59]. Given the cost of virtual reality headsets and related technologies, the rise in popularity of metaverse platforms may contribute to widening digital inequalities. Other sources of digital inequalities include older age, as older adults tend to be less autonomous and experienced in the use of new technologies [59], and physical disabilities, which to date have received little attention in the development of virtual reality hardware [60]. Digital inequalities are important to consider in the design of policies and the organization of services relevant to mental health care: these inequalities may translate into unequal access to digital mental health interventions, as well to health information and economic opportunities—all of which are determinants of mental health and recovery.

Virtual worlds are also a space where people can be exposed to bullying and discrimination. Exposure to cyberbullying is relatively common among the youth and has been associated with higher subsequent risk of psychological distress, suicidal ideations, and delinquency [61]. People in virtual worlds may feel emboldened to enact discriminatory and aggressive behaviors, as suggested by reports of sexual harassment, racism, homophobia, and transphobia within some video game communities [59,62,63]. Considering that 3D embodiment on an avatar, and its interactions with a virtual environment, can make internet-based interactions particularly vivid, one question that appears relevant for future research is to consider whether the metaverse risks magnifying the psychological impacts of internet-based bullying and discrimination.

People’s offline environment may further influence how they engage with virtual worlds and their propensity for using them in a problematic, addiction-like manner. Socioeconomic disparities and other forms of social adversity could increase people’s propensity for problematic metaverse use, for example, by motivating them to escape their stressful environment, or by constraining their access to alternative activities in their neighborhoods. Nagata et al [64] proposed that these effects of social adversity may explain, to some extent, why higher levels of problematic video game use are reported by adolescents in the United States who identify as Native American, Black, or Latinx, and by those with lower parental educational attainment.
Conversely, positive social interactions in the offline environment could be protective against the risk of problematic metaverse use. Illustrating this, a study of 250 MMORPG players found that higher levels of cultural consonance (ie, feeling successful in conventional society) and playing with offline friends were both cross-sectionally associated with lower levels of problematic gaming [38,65]. Similarly, in 2 samples of adult video game players, higher satisfaction of psychological needs in the offline world was cross-sectionally associated with lower levels of internet gaming disorder [66]. Young people’s interactions with their parents may be particularly important in shaping how they engage with metaverse technologies. For example, a community-based study of 2974 children and adolescents in Singapore showed that higher parent-child closeness was subsequently associated with a decrease in their levels of pathological video gaming after 1 year [67]. Interactions with friends, family, and society in the offline world are thus relevant to understand individuals’ engagement in virtual worlds. Together, these findings illustrate how individual and contextual factors, spanning dispositional, developmental, and social dimensions, may influence the mental health impacts of engagement in the metaverse.

**Conclusions: Immersion in the Metaverse**

When considered critically, the last 2 decades of research do not demonstrate universal harms of digital media on the mental health of the general population. The state of knowledge reveals many ways in which digital media may benefit or harm a person’s well-being, whether as a function of technological, individual, or contextual factors. But popular concerns persist and will almost certainly grow if the metaverse fulfills its commercial and social ambitions. Perhaps what sets the metaverse apart from previous technologies is the greater immersivity it affords, compared with traditional social media and video game devices. Immersion seems to be a factor that could amplify many of the mental health impacts described above: with greater immersion may come greater displacement of bodily awareness, greater embodiment over one’s avatar, greater sense of copresence with internet-based friends, and generally more vivid experiences. In turn, these features and experiences are likely to shape the motivations, affordances, and mental health risks associated with metaverse use.

It is too early to indicate whether the metaverse will broadly be a greater risk or benefit to mental health than previous digital media. Undoubtedly, virtual worlds are unprecedented in the scope of experiences and behaviors they can afford, and as such, in their potential to take a greater place in our daily lives. To adequately counsel patients and guide policies around the development and implementation of the metaverse, there is a need for timely and nuanced research on its opportunities and risks for mental health [68,69]. Arguably, previous research on digital media has often failed to achieve timely and nuanced knowledge production, partly due to the lack of clear theoretical frameworks and robust empirical methods [9,20]. Following these lessons, we believe a starting point is to examine the experiences and behaviors that are afforded and displaced by the metaverse, as well as the interplay of these effects with dispositional, developmental, and social factors. Collaboration among clinicians, researchers, individuals with lived experience, technology users, the industry, and other stakeholders will be crucial to successfully generate and translate this new body of knowledge. While metaverse technologies are in the process of being scaffolded and disseminated, psychiatry has a window of opportunity to think through conceptually and to strategically examine their risks and benefits for mental health.

**Acknowledgments**

VP receives grants from the Stratas Foundation and the Fédération des médecins résidents du Québec. SR receives a salary award from the Fonds de Recherche de Québec Santé and receives grants from Mitacs. HS has a Canadian Institutes of Health Research fellowship award, Mitacs fellowship award, and AGE-WELL award.

**Conflicts of Interest**

SR reports receiving a grant from Mitacs, is a steering committee member for AbbVie and is a shareholder at Aifred Health.

**References**

2. Stanton R. If Roblox’s daily users were a country, it would be bigger than Canada. PC Gamer. 2022. URL: [https://www.pcgamer.com/if-robloxs-daily-users-were-a-country-it-would-be-bigger-than-canada/](https://www.pcgamer.com/if-robloxs-daily-users-were-a-country-it-would-be-bigger-than-canada/) [accessed 2022-05-09]

https://games.jmir.org/2023/1/e43388

JMIR Serious Games 2023 | vol. 11 | e43388 | p.517

(page number not for citation purposes)


Abbreviations

MMORPG: massively multiplayer online role-playing game
A Conceptual Framework (2D-ME) for Explaining Self–first and Self–third Person Views of Prototyping Dynamics in Serious Games Design: Experimental Case Study

Sofia Hadjileontiadou¹*, PhD; Sofia B Dias²*, PhD; Leontios Hadjileontiadis³*, PhD

¹Department of Primary Education, Democritus University of Thrace, Alexandroupolis, Greece
²Interdisciplinary Centre for the Study of Human Performance, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal
³Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

*all authors contributed equally

Corresponding Author:
Sofia Hadjileontiadou, PhD
Department of Primary Education
Democritus University of Thrace
Alexandroupolis-Komotini Rd./New Chili
Alexandroupolis, 68131
Greece
Phone: 30 2551030103
Email: schatzil@eled.duth.gr

Abstract

Background: Design dynamics that evolve during a designer’s prototyping process encapsulate important insights about the way the designer is using his or her knowledge, creativity, and reflective thinking. Nevertheless, the capturing of such dynamics is not always an easy task, as they are built through alternations between the self–first and self–third person views.

Objective: This study aimed at introducing a conceptual framework, namely 2D-ME, to provide an explainable domain that could express the dynamics across the design timeline during a prototyping process of serious games.

Methods: Within the 2D-ME framework, the Technological-Pedagogical-Content Knowledge (TPACK), its adaptation to the serious games (TPACK-Game), and the activity theory frameworks were combined to produce dynamic constructs that incorporate self–first and self–third person extension of the TPACK-Game to Games TPACK, rules, division of labor, and object. The dynamic interplay between such constructs was used as an adaptation engine within an optimization prototype process, so each sequential version of the latter could converge to the designer’s initial idea of the serious game. Moreover, higher-order thinking is scaffolded with the internal Activity Interview Script proposed in this paper.

Results: An experimental case study of the application of the 2D-ME conceptual framework in the design of a light reflection game was showcased, revealing all the designer’s dynamics, both from internal (via a diary) and external (via the prototype version) views. The findings of this case study exemplified the convergence of the prototyping process to an optimized output, by minimizing the mean square error between the conceptual (initial and updated) idea of the prototype, following explainable and tangible constructs within the 2D-ME framework.

Conclusions: The generic structure of the proposed 2D-ME framework allows its transferability to various levels of expertise in serious games mastering, and it is used both for the designer’s process exploration and training of the novice ones.

(JMIR Serious Games 2023;11:e41824) doi:10.2196/41824

KEYWORDS
2D-ME conceptual framework; Technological-Pedagogical-Content Knowledge; TPACK; activity theory; AT; self–first and self–third person; Games TPACK; GTPACK; internal Activity Interview Script; iAIS; serious games prototyping
**Introduction**

**Background**

Serious games are games that focus on learning while entertaining. As their design entails elements from both the game and learning design, it is a process of a rather complicated thread of decisions. Conceptualizations of this process contribute to serious game design modeling frameworks [1]. With regard to serious game designing, the work of Hunicke et al [2] proposed the Mechanics, Dynamics, and Aesthetics (MDA) framework. In particular, their work is fundamental in the area and realizes 3 distinct components that can be viewed from the user’s and the designer’s perspective. The Mechanics component describe specific options of actions that the user can perform while interacting with a game, and they can be described through verbs (eg, move). On the other hand, the Dynamics component refers to the combined player options toward a system of interactions at a higher level than the Mechanics component, that is, they can be considered the system-behavior of a game. However, the Mechanics component, being more specifically defined at a lower level of abstraction, is more reliable than the Dynamics component. Finally, direct and indirect interactions of the player with the game result in the Aesthetics component that refers to the emotional responses of the player (eg, fun), as they are provoked by the combination of the game Mechanics and Dynamics components, the latter bridging the Mechanics component with the Aesthetics component. Thus, a flow is realized during the game design, that is, an initial selection of the Mechanics component, then the Dynamics component, and finally the Aesthetics component. Although the work of Hunicke et al [2] was criticized as situational [3], it can serve as a direct framework to outline the serious game design. A detailed list of core mechanics was proposed by Järvinen [4] and a list of comprehensive Dynamics was proposed by Pendleton and Okolica [5].

Research in the area of serious game designing resulted in further frameworks that followed the work of Järvinen [4]. Moreover, Ávila-Pesántez at al [6] provided a systematic literature review regarding the methodologies, frameworks, and models applied to game designs in the period from 2008 to 2016, resulting in 11 approaches, whereas Viudes-Carbonell et al [7] provide a state of the art of 6 models in the area. However, the Learning Mechanics–Game Mechanics Framework [8] is worth mentioning, as it was based on the idea that learning occurs while interacting with the game, that is, Learning Mechanics can be mapped to the game’s Dynamics component and relevant thinking skills according to the revised Bloom taxonomy [9]. Thus, the learning objectives are diffused in the serious game design [10] to entail the game’s educational character. Following this approach, Pendleton and Okolica [5] proposed the Game Design Matrix structure, which reorients the building block of the MDA framework [2] by first selecting the Dynamics component, according to the expected level of mastery in the Bloom taxonomy, so that the emphasis is initially given to the learning objectives that drive the learning outcome of the serious game. However, the design procedure is an iterative process and Hunicke et al [2] underline the importance of iterative analyses and refinement of the game design results.

In a more detailed approach, Viudes-Carbonell et al [7], following the MDA framework, consider the game design as iterative cycles of design, test, evaluation, and redesign and stress the idea that the iterations would eliminate the risk of failure while designing a serious game. In particular, they focus on the prototyping iterations and consider them as series of small steps toward the enhancement of the quality of the game outcome. Following this path, Fullerton et al [11] consider prototyping an important part of the game design. Prototyping may take place by various means, for example, by pencil and paper, computer, and other artifacts in many disciplines (eg, engineering), to break out a complex problem to subproblems and work on these separately. However, a game is rule based; thus, it is sensitive to changes and/or prediction of possible impacts owing to rule changes. In this perspective, prototyping may resolve unpredictability issues in game design [12]. Moreover, considering games as pieces of art when compared with average software [13], their designing is even unpredictable as the design space may dynamically evolve through the emergence of new ideas and creativity.

The aforementioned approaches provide the frameworks of designing and establish steps for the enactment of design ideas. However, they do not refer to the knowledge that is needed for the materialization of the game design. With regard to this direction, the frameworks described in subsequent sections have been considered.

**Design Frameworks and Serious Games**

**The Technological-Pedagogical-Content Knowledge Framework**

The Technological-Pedagogical-Content Knowledge (TPACK) was proposed as a framework to account for the knowledge needed by the teachers to use information and communications technology in their classrooms [14-16]. In particular, this framework extended the notion of Shulman [17] concerning the pedagogical content knowledge (PCK), by foreseeing the inclusion of knowledge related to technology. Figure 1A depicts the TPACK framework and manifests the interplay between forms of knowledge (content knowledge [CK], pedagogical knowledge [PK], technological knowledge [TK], PCK, technological PK [TPK], and technological CK [TCK]). More specifically, CK refers to the knowledge of the content of the lesson, for example, the concepts that are to be taught; PK refers to the pedagogical considerations as to how to teach the content; TK refers to the knowledge about the technological affordances and their use; TCK and TPK refer to the way the technology could support and materialize the delivery of content and pedagogy; and finally, PCK refers to the pedagogical way in which the content could be delivered to the learners. The contextualized synthesis of all these forms of knowledge constitutes the core TPACK framework. TPACK is usually presented as 3 overlapping equal cycles denoting 3 forms of knowledge (CK, PK, and TK) and their overlaps (TCK, TPK, PCK, and TPACK), all included in a circle that denotes the knowledge that is needed for the integration of technology in a relevant sociocultural context. TPACK has been widely used as a framework to reflect information and communications
The TPACK-Game Framework

TPACK serves as a theoretical reference framework of necessary knowledge for the technology integration in different contexts; for example, Foster et al [19] proposed TPACK as an aid for the teacher when selecting a game to realize its constraints and affordances for technology, pedagogy, and content. However, Willermark [20] commented on TPACK for its abstract reference to general technology that could be used in the classroom, that is, it lacks the specificity that might contribute to further specification of knowledge that is needed. In this regard, in the specific case of computer game integration in the classroom, Hsu et al [21] proposed TPACK-Game (TPACK-G; Figure 1B) to reflect the knowledge that the teachers need in this case. In particular, they define game knowledge (GK) as the knowledge about the general use of computer games, game PK (GPK) as the pedagogy related to the way a game is used in the classroom, and game PCK (GPCK) as the “knowledge of using games to implement teaching methods for any targeted content” [21]. On the basis of prior work on the developmental pathway of TK and TPK to TPACK, Hsu et al [21] proposed the GPCK as a pyramid, where GK constitutes the basis, followed by the GPK, and finally, at the top, the GPCK. Upon this modeling, they claim that for the development of specific GPCK, the GK and GPK are prerequisites, so a lack of the foundation knowledge of how to play a game, that is, familiarity with gaming environment (GK), leads to a lack of GPK and therefore GPCK [21]. Structural relationships were also detected between the TPACK-G attitudes toward games and actual teaching use [22].

Both the TPACK and TPACK-G frameworks are proposed through a “stationary-like” approach, whereas a more dynamic one might reveal the evolution of types of knowledge over time while an activity takes place, for example, during the iterative approach toward the designer’s activities while designing a serious game. The activity theory (AT) framework may provide a lens toward this direction, as explored in the subsequent section.

The AT Framework

Overview

The AT framework has its roots in the 1920s and provides a conceptual framework according to which, the human activity connects the individual world of the subject with the social one (Figure 2). Upon its initial formulation [23, 24], the AT framework focused on the activities of the subject, for example, human being. For the subject to meet his/her needs, he/she performs activities by interacting with objects in the world by means of tools that mediate the interaction. Thus, the subject is involved in an activity motivated by an expected outcome (the “why” of the activity), pursues goals that direct the activity (the “what” of the activity), and performs operations (the “how” of the activity) [25]. However, the connection between subjects and activities is bidirectional as they mutually determine one another. Basic principles of AT have evolved through the works of Leontiev [23, 24], namely [26] (1) object orientedness, which states that objects differentiate the activities that are directed to them upon motivation toward a need; (2) hierarchical structure of activity, which is a hierarchical system that shows the dynamic relationship among 3 levels of human activity. From a top-down perspective of the activity, at the top, the activity driven by the motive is considered. However, the activity in a lower level is performed through conscious actions that are directed to goals (possibly decomposed to subgoals) that are to be undertaken to fulfill the object. At an even lower level, actions are performed through unconscious operations that are orientated to conditions under which the goal is to be reached; (3) tool mediation, toward a purposeful activity; (4) development, as a context to analyze the activity, for example, from the research respective, this leads to the analysis of the dynamics of the object transformations over time; and (5) internalization externalization.
Synergies of AT in Serious Games Prototyping

In the work of Carvalho et al. [27], the AT framework was used for the analysis and design of serious games. In particular, they proposed the AT-based conceptual model of serious games, which contributes to the realization of the components of a game and their roles and to the recognition of the educational objectives. In addition, they foresee 3 activities: the gaming, the learning, and the instructional activity and their tensions to describe the contexts in which the game is used. Moreover, for each activity, they propose elements taxonomized into actions, tools, and goals per activity, upon which a game may be analyzed or designed. However, the AT-based model of serious games is restricted only at interactions between the triad, that is, subject, object, and tools. The works of Viudes-Carbonell et al. [7] and Manker et al. [28] use the AT approach to propose analyses of the prototyping procedure during the design of a serious game. In particular, Viudes-Carbonell et al. [7] focus on the iterative character of the serious game design and present an early-stage methodology and a prototype for learning a case-concept. Moreover, Manker et al. [28] also focus on the game prototypes and among other, on the role of the prototype to the internalization and externalization of the designer’s ideas. The combination of the AT framework with the TPACK framework, namely TPACKtivity, has been proposed by Terpstra [29] as a lens to follow preservice teachers’ (PTs’) TPACK development through their activities, that is, realizing its dynamic character across the time.

Contradictions, Social and Relational Self-views Within AT

The traditional logic considers contradictions in a system as problems; dialectical logic, however, considers that learning and development are driven by them. Hence, within the AT framework (Figure 2), analysis of the trajectory of the object across time can contribute to the realization of the evolution of its current existence and the expected contradictions that may drive its further development [26]. Thus, a deep assumption is that the activity develops across the time mainly through contradictions that arise, that is, upon resolving 1 contradiction, another may arise. Continuous approach of the activity provides continuity in viewing the driving forces (triggers and resolutions of contradictions) that steer this activity further. Focusing on the subject of the activity system, initial ideas about the dichotomy of the context-free information processing of human functioning as opposed to society were further elaborated, extending the subject approach in the AT. In this direction, Stetsenko and Arievitch [30] proposed a framework to combine the social and relational view of the self. In their approach, the subjectivity that is developed outside and in the activity system of reference, may result from the role according to the social position of the human being (ie, serious game designer), which along with social interactions and collective practices coevolve. On the other hand, the self is considered as an active agentive role that an individual does not only acquire historical cultural norms and experiences but also develops them further through change and novelty. In this sense, the ever-expanding social practices entail human subjectivity, and both social and relational views of the self, coexist, or self-coexist among planes of the activity through the internalization and externalization processes.

The aforementioned design frameworks present either the design of a serious game through the lens of the AT framework or the combination of TPACK with the AT framework. However, they hardly focus on the subject of the activity (ie, the designer), and when they do so, a macrolevel is adopted, referring mostly to the procedures per se, for example, the connection of the subject’s internal plane with the world’s plane.

Study Aims

In this paper, we introduce the 2D-ME framework, an innovative conceptual framework that adopts a microlevel approach toward the serious game designer’s inner world while he/she is prototyping. In particular, we dynamically followed the designer’s first- and third-person views (2D-ME) during the prototyping activity under the lens of the AT framework and an extended version of TPACK in the game design area, namely Games TPACK (GTPACK). Under this aim, the following research questions were investigated:

1. Is 2D-ME a useful framework for describing the prototyping dynamics during the design of a serious game?
2. How do the internal (first person) and external (third person) views facilitate the optimization of the prototyping output during a serious game design?

Although we refer to the serious game designer, the 2D-ME framework could easily be transferred to the case of serious game developer (sometimes these 2 roles are undertaken by the same person). Moreover, as it offers new insights in the internal processes during prototyping, the 2D-ME framework could also
be used as a research framework for analyzing the research aspects of dynamics during creative prototyping.

Despite the conceptual character of the proposed 2D-ME framework, its practical implementation can easily be realized via a series of steps:

1. Conceptualization of the initial idea target of the serious game
2. Identification of the AT constructs
3. Activation of a metacognitive process (e.g., reporting in a diary via a self-interview script), which could facilitate the interplay between the roles of first- and third-person view during the prototyping
4. Enabling and capturing of the dynamics between the AT constructs
5. Adaptation of the AT dynamics for the convergence of prototype successive versions to the initial idea target of the serious game

**Methods**

**The Proposed 2D-ME Conceptual Framework**

**Main Constructs and Dynamics**

Stemming from the design frameworks presented so far, the proposed 2D-ME conceptual framework is described in this section. Figure 3 depicts the main constructs of the 2D-ME framework. As it is apparent from the comparison between the shapes of Figure 2 (AT) and Figure 3 (2D-ME), a connection of the 2D-ME framework with the TPACK, TPACK-G, and AT is noticeable. This is materialized by the inclusion within the 2D-ME framework of (1) the concept of triangles and interactions as in AT and (2) the combination of TPACK with the TPACK-G as a new entity in the AT triangle, namely GTPACK. However, all constructs of AT (Figure 2) are replaced with new ones (Figure 3), as follows:

- Instruments or mediating artifacts → GTPACK(t); t = 1, 2, ..., N
- Subject → Subject S(i,t), i = first-person or third-person view; j ≠ i
- Rules → Rules R(t); t = 1, 2, ..., N
- Object → Object O(t); t = 1, 2, ..., N
- Division of Labor → Division of Labor DL(t)
- Community → Subject S(j,t), j = first-person or third-person view; j ≠ i
- Outcome → prototype versions P(n,k), n = 1, 2, ..., N; k = 1, 2, ..., M

As it is apparent, the 2D-ME framework embeds within its structure a dynamic activity within its constructs that interact and (potentially) are modified across time (t). In particular, the prototype designer is considered as the subject S(i,t) and subject S(j,t), j ≠ i, alternating across time (t) between the first-person and the third-person views. The S(i,t), motivated toward the materialization and optimization of the object O(t), that is, the game prototype P(n,k), n = 1, 2, ..., N; k = 1, 2, ..., M, at specific time instances k, uses their current GTPAK(t) to produce each prototype version (P(:,1), P(:,2),...) toward P(:,M), which is the final version of the outcome of the activity system. It should be noted that the n parameter expresses the different dimensions of prototype (see The Object Outcome [Prototype] section). Apparently, the first-person and the third-person view of the subject communicate upon rules R(t) and division of labor DL(t).

As it can be deduced from the aforementioned 2D-ME structure (Figure 3), the flexibility of the AT framework to project the activity at different levels of analysis is used here to shift from a macro- to a microscale approach. More specifically, at the macroscale approach, the prototyping procedure in the game design provides samples of the evolution of the game design across the time. From a main concept perspective, the 2D-ME framework adopts a microscale approach by following the inner iterative prototyping procedure that follows the game designer across the time. Upon this, an AT triangle is defined, in which the subject, that is, the designer of the serious game, is the focus of the approach. The 2D-ME framework foresee the designer to adopt 2 perspectives while prototyping, that is, the first-person and third-person views. In this way, 2D-ME follows the internalization and externalization actions within the game designer and provides a conceptualization of the prototype design before its externalization to the others (collaborators, users, etc). To further explain the main concept, a description of each entity of 2D-ME is presented in subsequent sections.

**Figure 3.** Schematic representation of the main constructs that form the 2D-ME framework. DL: division of labor; GTPACK(t): Games Technological-Pedagogical-Content Knowledge (dynamic); O: object; P: prototype; R: rules; S: subject.
The GTPACK

The GTPACK (Figure 1C) is considered in 2D-ME as a tool that extends the TPACK in the area of the necessary knowledge base for the serious game designing. In particular, the Game CK refers to the cognitive content of the game, that is, the subject matter of the discipline of reference. Game Technological Knowledge refers to the knowledge base relevant to the technology related to the games design and its use. Finally, Game PK reflects the knowledge of possible misconceptions, learning theories and how to apply them in games, metacognition, etc. On the basis of this perspective, in a game designing procedure, Game Technological CK might be triggered for the representation of the concept to be learned in the game environment, the GPCK might be triggered for the definition of the game dynamics, and the GTPK might be triggered for the selection of the game mechanics, whereas the GTPACK might be triggered for the whole game design and aesthetics. The GTPACK serves as a tool that mediates the activity of the subject upon the object toward the outcome. It should be noted that the GTPACK is not a stable framework across design time, as the designer might interact with resources (either human and/or material) that may alter its load and dynamics across the time.

The Subject

The 2D-ME framework proposes 2 points of view of the subject, the first-person and the third-person view, considering transitions from the internalization to externalization processes and vice versa, yet both defined within the same person. In particular, the first-person view considers the subjective version of the subject in the activity system that reflects his/her activity following rules of designing a serious game, that is, the social world. The third-person view considers the self-version of the subject that reflects his/her activity according to the object of the activity system that he/she acts upon as the designer. The third-person view is the agentic self who leads the activity as he/she is engaged with the social world. On the basis of this perspective, the first- and third-person views hold a GTPACK and they both coevolve with the activity, yet the third-person view is mainly reflective, to actively introduce possibilities; enhance them by prioritizing other actions; and possibly generate activities that stem from the reality, fantasy, and serendipity.

The Rules

In the AT framework, the rules primarily mediate how the subject acts in relation to the object, including the tools used and the ways they are used [31,32]. This could include specific and/or well-established patterns of behaviors that could be followed because of either cultural norms and/or other reasons (eg, professional and legal mandates). In 2D-ME, the rules refer to the ways of communication between the first- and the third-person views during the subject’s communication. Apparently, this interpersonal communication follows some norms (rules) that facilitate the process by which the subject is engaged in unspoken internal dialogue among different and sometimes conflicting attitudes, thoughts, and feelings, all represented by the first- and third-person views. Clearly, these rules are subject dependent and can be imposed, sustained, redefined, modified, or dropped within a dynamic process that could add to the metacognitive skills of the subject.

The Division of Labor

The main responsibilities (what is being done by whom) that are involved toward the object define the division of labor in AT, also taking into consideration the horizontal division, that is, across tasks, and the vertical division, that is, across the power, positions, access to resources, and rewards [31,32]. In 2D-ME, the division of labor refers to the contribution of 2 subject’s views, that is, the first- and third-person views, toward the object within the interpersonal communication. The concept of horizontal and vertical division can also be transferred at this inner level by considering the different aspects that could be triggered during the interpersonal communication. For example, the horizontal division could include a systematic definition of a sequence of interpersonal sessions across the prototyping process. Moreover, the vertical division could prioritize the first-person view in the access to and use of technological resources (eg, game developing software), leaving space for the third-person view for reviewing the aesthetics (eg, rewarding and/or corrective feedback on game sensation, fantasy, or challenges).

The Object Outcome (Prototype)

The object in AT can be approached by various views, that is, the object is seen as a thing to be acted upon, an objectified motive, or a desired outcome [33]. The object in 2D-ME is to materialize an idea toward optimizing the outcome, which is a prototype of the game. The prototypes could vary in purpose, as they can refer to a role, an implementation approach, a look, or a sense of feeling. These include interaction with the users (role), the construction of the game (implementation), and users’ experience (look and feel) [21]. This multiplicity in roles results in multiple dimensions (denoted with the parameter n, with N being the total number of dimensions). The various versions of the prototype in 2D-ME (ie, P(n,k), n = 1,2,...,N; k = 1,2,...,M) are tangible externalizations of the development transformations that the activities undergo between the constructs across the serious game design process.

Prototyping Optimization

The proposed 2D-ME conceptual framework considers the prototyping process as an optimization process that converges to the optimal prototype (P(N,M)) by minimizing a cost function. This is clearly depicted in Figure 4, where the constructs triangle (Figure 3) is used as an adaptation engine. The cost function per k is the convex function of the mean square error (MSE) between the conceptual (initial and updated) idea of the prototype and its current version (P(n,k)) is presented as follows:

The adaptation engine provides the convergence, so the whole activity gradually reaches to the optimized outcome, that is, P(N,M). However, the optimization convergence is not seen within an absolute perspective but rather from the designer’s view and experience. This means that the current optimization
process could stop at a point that the \( P(N,M) \) will not express the full spectrum of the designer’s skills and knowledge, but the process will reach at a point where the designer has already concluded for the validity of the final output and the sufficient use of their knowledge and experience, that is, their current level of GTPACK.

Furthermore, as it can be seen from Figure 4, the initial idea \( P_0(n); n = 1,2,...,N \) acts as a trigger to the adaptation engine and, at the beginning of the process, it is considered as static. However, the processes within the adaptation engine (eg, the interpersonal communication), apart from the update of the version of \( P(n,k) \), could potentially affect the initial idea, as well. In this vein, its transformation, that is, \( \mathcal{C} \), is considered at the next stages of the design process. This transformation could alter specific dimensions of the prototype (eg, eliminate some and add new ones) and/or the characteristics of each dimension. Apparently, the frequency of the changes of the \( \mathcal{C} \) is anticipated to be low, considering the level of the designer, that is, how clear and solid is the initial idea in his or her mind. To this end, during the optimization process, some versions of the \( \mathcal{C} \) could be constant across \( k \), for example, \( \mathcal{C} \). This is, usually, anticipated as the prototype converges to its optimum version, because the designer has already a clear view of its intended and the constructed version.

**Figure 4.** Schematic representation of the prototyping optimization process in 2D-ME framework. The constructs triangle (Figure 3) is used as the adaptation engine toward the minimization of the difference of the conceptual (initial and updated) idea of the prototype \( P(n,k) \), that is, gradually reaching to the optimization of the prototype (at \( k=M \)) across its intermediate versions (at \( 1 \leq k \leq M \)) and across its dimensions. MSE: mean square error.

**Capturing of Dynamics**

From what it is presented so far, it is clear that 2D-ME entails dynamics among its constructs and in the prototyping optimization process. To capture such dynamics, some sampling tools should be used. Clearly, the interpersonal communication (between the self–first person and the self–third person) that is involved in 2D-ME sets the highest hurdle in the dynamics capturing process. One way to infer for the activation of such communication could be the use of a systematic approach that would include a subject’s diary. The latter could be structured on specific contextual factors reflected in a set of statements. A characteristic example of such statements is included in the activity checklist [34,35], which is based on the AT. As the activity checklist provides a “contextual design space” [28], this could be adapted, accordingly, to include statements that could help the designer to describe in his/her diary the dynamics used between the 2D-ME constructs. In this way, the activity checklist itself can serve as a “valuable aide memoir and a tool for reflexivity” [36]. Nevertheless, the activity checklist contains many items (43 for design and 37 for evaluation) [35] that make its full use quite difficult in practice. As a remedy, adaptation of the activity checklist to the form of interview questions, limited in number, was proposed by Duignan et al [37]. Hence, this interview form of activity checklist could trigger reflection and self-talk and could assist in capturing specific dynamics at the various phases of the prototyping process. In this vein, internal questions related with the 2D-ME dynamics that could be self-answered and documented in the diary could be formed, as an internal Activity Interview Script (iAIS), tabulated inTextbox 1.

An additional source of understanding the dynamics is via the exploration of the alterations across the different versions of the prototype. Apparently, this is a lower quality sampling of the dynamics when compared with the aforementioned approach; it provides, however, a way of monitoring the dynamics, especially when exploring the differences across the dimensions \( n \) of the prototype versions across the design time. A process of deconvolution is then applied to identify the constructs that were more frequently used in a current version of the prototype. Projecting this construct activation across the whole prototyping process, the dynamics per construct can be revealed.
Question 1. Goals-related questions

• Question 1.1. What are the different roles that you identify in yourself and are involved in the prototyping?
• Question 1.2. How do you breakdown, in a step-by-step form, your prototyping process?
• Question 1.3. How this fine-graining process can vary?
• Question 1.4. How do you know that you have successfully completed each intermediate goal?
• Question 1.5. How could you evaluate the achievement of your higher-level goals?

Question 2. Contradictions-related questions

• Question 2.1. What contradictions can you identify between your different roles?
• Question 2.2. How do you resolve such contradictions?

Question 3. Tools, transition, internalization, externalization–related questions

• Question 3.1. How does Games Technological-Pedagogical-Content Knowledge support the transition between the first-person and third-person view?
• Question 3.2. How does Games Technological-Pedagogical-Content Knowledge affect the way you think and reason about the prototyping activity?
• Question 3.3. How difficult is to perform the different interpersonal roles?
• Question 3.4. How do you use representations of your work between the different interpersonal roles?
• Question 3.5. How do you internally handle the complexity of the prototyping process?

Question 4. Rules, division of labor–related questions

• Question 4.1. How do explicit or implicit rules, norms, and procedures affect your different interpersonal roles?
• Question 4.2. How do you organize the different interpersonal roles across the prototyping process?

Experimental Case Study Setup

An experimental case study was set up as a running example of the realization of 2D-ME in practice. In particular, the case refers to a PT as the designer of a serious game for kindergarten through grade 12 students. The motivation for the latter was drawn from the physics student’s exercise textbook (page 63) for the fifth grade for primary school in Greece. In the latter, an exercise related to the light reflection on mirrors presents a situation in which a series of mirrors are fixed at different positions in a box. At the perimeter of the box there are 4 openings where specific objects (compass, scissor, and pencil) and a boy’s eye are positioned. The student is asked to identify and draw the paths of the light reflections on mirrors so as the boy can see the 3 objects (see $P_d(n)$ in Figure 6). This exercise was presented to the PT as a stimulus for the design of a related serious game on light reflection. The PT was at his final semester of his studies at the Department of Primary Education, Democritus University of Thrace, Greece, and he already had attended 1 semester concerning serious game design using Scratch [38]. In particular, concerning his GTPACK, the PT had Game CK about the light mirror reflection, which was the concept to be practiced by the serious game. Moreover, he had extended GPK because of his studies and good Game Technological Knowledge based on the familiarity with the game prototyping, design, and implementation in Scratch. Thus, his current GTPACK was considered adequate for this experimental study. It should be noted that the PT was already aware about the GTPACK framework.

The research design foresaw controlled experimental situation that took place at the Democritus University of Thrace computing laboratory within 1 month. The PT willingly participated in this experiment and was informed about the experimental process before embarking on it. More specifically, he was asked to express in a diary the conceptualization of his initial idea $P_d(n)$ of the serious game and then to proceed to his prototyping activity and optimization (Figures 4 and 5). Moreover, he was addressed with the aforementioned iAIS and was advised to document his relevant thoughts in the diary during the prototyping procedure. It was made clear to the PT that the use of the diary should be spontaneous and in no way should hamper his creative impetus. Furthermore, he was informed and consented that an observation documentation would be made by a researcher (first author, SH) to capture the explicit prototyping performance of the PT. At the end of the experiment, the PT consented to openly share the diary; a thematic analysis of the diary was manually conducted and combined with the data from the observations along with the outcome of the whole activity.
Figure 5. Schematic representation of the dynamics across the prototyping optimization for $k=1,2,...,5$. At the top, the Games Technological-Pedagogical-Content Knowledge (GTPACK) alteration across are displayed, where the dynamics affect its internal set (at $k=1,4$). The constructs dynamics are denoted with arrows, accompanied with the related internal Activity Interview Script (iAIS) questions (Textbox 1). The optimization process is depicted for each outputted prototype, whereas the estimated mean square error (MSE) is depicted at the bottom, showcasing its reduction as the game design evolves.

Figure 6. Illustration of the produced game prototype scenes, along with the addition of the Scratch programming code at each $k, k=1,2,...,5$. The initial idea, $P_0(n)$, that acts as the stimulus for the creation of the light reflection game is also depicted.

Results

Conceptualization of the Initial Idea

The PT formulated his/her initial idea $P_0(n)$ of the serious game, based upon his/her GTPACK. Textbox 2 shows what he wrote in the diary.

From the excerpts mentioned in Textbox 2, it can be inferred that the PT conceptualized his initial idea $P_0(n)$ of the serious game. He also defined the $N=6$ dimensions, where $n=1$ corresponds to the space (sprites: a source of the light ray, one mirror and one target), $n=2$ denotes the dynamics (hidden objective, player autonomy, limited actions, and feedback), $n=3$ refers to mechanics (move, rotate, aiming and shooting, and allocating), $n=4$ corresponds to the sprite costumes (a torch, a line, a top view of a mirror, and a top view of a brick), $n=5$ refers to the aesthetics, and $n=6$ is for the programming dimension.
I can extend the image in the school textbook where there is a source that produces the light ray and mirrors at stable positions to reflect it so as through successive reflections it can reach a specific point. The students are asked to hypothesize the reflections and then draw a line to depict the path of the light ray as it reflects in this labyrinth of mirrors. Thus, if the student realizes this path once, then there is no interest to resolve it. So, I need to extend this activity towards a more interactive and interesting situation through my serious game. Keeping the analogy with the textbook activity, where it is presupposed that the student already has been taught the light mirror reflection, my educational aim is to involve the students in a serious game, in order to further elaborate the light reflection concept in more dynamic situations than the textbook, eg, the student can try the mirror reflection towards different lines of the ray and not in a specific labyrinth. Thus, my objective is to move the cognitive effort from the applying level and beyond according to the Bloom’s taxonomy. Moreover, I can use depictions from the textbook that the students are acquainted of, i.e., a torch as a source for the light ray, a line for the light ray, top view of a mirror and a top view of a brick as target.

I will elaborate on the well-known K-12 students’ misconceptions about the light reflection, concerning the equality of the angles of the light ray as it lands on and reflects from the mirror surface. I will assume that the students had already been taught the light mirror reflection; thus, they have conquered the first two levels of the Bloom’s taxonomy, so my serious game will not provide relevant information.

My objective is to move the cognitive effort concerning the aforementioned concept at the applying level of in the Bloom’s taxonomy.

As a tool I will use Scratch.

I will follow the GDM design framework and I will choose dynamics and mechanics from those presented in the work of [5], combined with additional important elements, namely:

- **Space and Dynamics**: The student should manage to apply the idea of the equal angles that are employed in the mirror reflection, but I do not want to show this; they must hypothesize the light line in a space arrangement that will include three sprites, ie, a source of the light ray, one mirror to reflect upon and one target where the reflected ray should hit, respectively. Thus, I choose the hidden objective dynamic. I also want the student to arrange in a way the game space, so I choose the player autonomy dynamic; yet, with limited actions so as not to be confused. Finally, I will provide feedback to his/her efforts, eg, a score counting the successive efforts toward the shooting the target with the reflected light ray.

- **Mechanics**: The torch will be at a stable position, but the student could rotate it. The mirror would be horizontally positioned in the bottom of the scene and could be moved in both directions on the x axis and the target will be at a stable position. Thus, I choose move, rotate, for the objects, aiming and shooting for the light ray, allocating for the hypothesized path of the light ray.

- **Sprite costumes**: I will use depictions from the textbook that the students are acquainted with, ie, a torch as a source for the light ray, a line for the light ray, top view of a mirror and a top view of a brick as target.

- **Aesthetics**: I think that the acquaintance of the students with the sprite costumes, the appealing scene, the explicit roles of the sprites and the challenge to target the brick will motivate the students to have fun and keep playing it toward higher scores. All these will contribute to the esthetics dimensions of my serious game, ie, its sensation (sense of pleasure), fantasy (make-believe), challenge (obstacle course), discovery (uncharted territory), and expression (self-discovery).

- **Programming**: I need to program all the needful with Scratch.

### Dynamics in Adaptation Engine and Prototype Optimization

#### Overview

During the activity, the PT followed iteration paths through which a series of enhancement of the prototypes that were developed in Scratch resulted in the final output (the game). Here, indicative versions of the prototype are presented to reflect the dynamics and realization of the adopted optimization procedure within the 2D-ME framework. In particular, we followed the $P(n,k)$, $n = 1,2,...,6$, $k = 1,2,...,5$, prototypes in the sequential order that they were developed, cumulatively constructing the serious game aspects. It should be noted that not all dimensions are simultaneously activated across the prototyping, because the PT placed the effort according to his temporal preferences and focus on specific dimensions.

#### Diary-Based Sampling

As a first means of dynamics acquisition, excerpts from the PT’s diary were used for each version of the prototype. Figure 5 facilitates the presentation at multiple levels that are used during the application of the 2D-ME framework, showcasing the activity system that the PT was involved in, its dynamics across the triangles’ vertices (Figures 3 and 5: $\{A,B,C\},\{a,b,c\}$), and the optimization procedure that took place. The corresponding activated dimensions (separated by semicolon) are described below; researcher SH’s external view based on the iAIS is also noted within brackets and included in Figure 5:

- **Initial idea**: $\text{Prototype: } P (1;6,1)$
  - a→A: “From the first person view, I will start by setting the main scene and creating the first sprite; however, my GTPACK is not adequate to achieve this” (Question 1.2). “I need to enhance my GTK on how to program the reflection of the light ray on a moving mirror” (Question 3.2).
  - A→C: “I need to incorporate multiple roles within myself to activate a reflective path for my design decisions” (Question 1.1 and Question 4.2).
  - C→a: “From the first person view, I now feel more confident to start materializing my $P(n)$ idea” (Question 3.2).
  - a→c: “From the first person view, I managed to write the Scratch code for the main scene and the first sprite and produce the first version of my object, ie, $P(1;6,1)$” (Question 1.4).
• Initial idea: Prototype: P(1; 1; 4; 6, 2)
  • a→b: “From the first person view, I will next put the mirror in the scene” (Question 1.2). “From the third person view, I realized that I need to revise my initial idea for the size and the way the mirror should look like, to avoid trivial game options” (Question 2.1).
  • b→B: “The third person view helps me to identify design pitfalls and use one rule, i.e., to be on the shoes of the user, in order to better understand the game” (Question 4.1 and Question 4.2).
  • B→a: “From the first person view, I keep as a rule to really play the prototype, in order to feel the pitfall that I conceived as such from the third person view” (Question 1.1 and Question 4.2). “I also follow the rule to distinguish between what remains the same and what should be changed towards optimization” (Question 1.4).
  • a→c: “From the first person view, I understand that the third person view scaffolds my thoughts for the next step towards the implementation of my initial idea and I managed to write the Scratch code for the mirror, selecting the analogous size and view, producing the second version of my object, i.e., P(1; 4; 6, 2)” (Question 1.4 and Question 3.4).

• Initial idea: Prototype: P(1; 3; 4; 6, 3)
  • a→b: “From the first person view, I will next put the brick and the light ray in the scene and adopt the sprite costumes of a line and top view of a brick, respectively” (Question 1.2). “From the third person view, I realized that I need to revise my initial idea for the fixed angle of the light source to avoid trivial game options” (Question 2.1).
  • b→C: “I realized that I need to have a more organized sequence of reflections; hence, I have decided to adopt more frequent alternation between the first and third person view during my design process” (Question 4.2).
  • C→a: “From the first person view, I now feel more excited to put more degrees of freedom to the torch” (Question 1.3).
  • a→c: “From the 1st person view, I managed to write the Scratch code for the addition of the brick, the light ray and the rotation of the torch, producing the third version of my object, i.e., P(1; 3; 4; 6, 3)” (Question 1.4).

• Initial idea: Prototype: P(2; 3; 6, 4)
  • a→b: “From the first person view, I will next include the feedback to the student via a score display; however, I feel that something is missing...” (Question 1.2 and Question 2.1). “From the third person view, I had the idea of inserting randomness in the game. I find myself a bit strange when I have ideas that I do not know how to implement. It is a contradiction to my general attitude to proceed safely. However, when I follow them, I believe that new possibilities are revealed.” (Question 1.3, Question 2.1, and Question 2.2).
  • b→A: “I feel that I need to expand my GTPK towards the motivational and engaging role of the surprises in the game through randomness and its relevant programming” (Question 3.2).
  • A→C: “I can use my enhanced GTPACK and identify better the involvement of my different roles in myself for the technology- and aesthetics-related aspects of the prototyping” (Question 3.1, Question 4.1, and Question 4.2).
  • C→a: “From the first person view, I broke down the move mechanic across my prototyping procedure, in order to realize which objects it should refer to, and materialize the decided randomness” (Question 1.2, Question 1.3, and Question 3.5).
  • a→c: “From the first person view, I managed to write the Scratch code for the addition of the score and incorporate movement of the sprites, producing the fourth version of my object, i.e., P(2; 3; 6, 4)” (Question 1.4).

Initial idea: Prototype: P(5; 6, 5)
  • a→b: “From the first person view, I feel that, most probably, I have reached the point of satisfaction with the structure of the game so far; but I am not 100% sure...” (Question 2.1). “From the third person view, I visualized a more interesting approach, focusing on the game aesthetics, that incorporated the sense of depth in the main scene” (Question 2.2 and Question 3.4).
  • b→a: “From the 1st person view, I understood that this reflection provides me with many options to increase the sensation and fantasy of the game, using different 3D scene backgrounds” (Question 3.4).
  • a→c: “From the first person view, I managed to write the Scratch code for the addition of a 3D background in the scene, producing the fifth version of my object, i.e., P(5; 6, 5)” (Question 1.4). “I believe that, at this point, I have achieved my first higher goal to construct a good prototype of these functionalities and aesthetics of the game, converging to the initial idea conceptualized in the form of P(5; 6, 5)” (Question 1.4 and Question 1.5).

It should be noted that in the excerpt of the optimization process presented earlier, the initial idea was kept constant across k, as the level of the game designing was of low complexity, trying to conceive and materialize the basic environment and functionalities of the game. As the game advanced, updates of the P(n) took place, for example, through the insertion of >1 mirrors spreading randomly in the scene space and introduction of various game levels with different degrees of difficulty (e.g., random distribution of nonreflecting obstacles and time limitation per level). Moreover, the GTPACK of the PT was enriched during the prototyping process, showcasing the internal dynamics that are developed across the design time at the GTPACK level, depicted as expanded overlapping spaces in Figure 5 (top). The estimation of the MSE (Figure 5, bottom) was based on the comparison (equation 1) between the current and the P(n;k) across the n dimensions. As the dimension of
programing ($n=6$) is not included in the MSE calculation, involving the dimensions of $n=1,2,\ldots,5$ in equation 1. As at each $k$, not all dimensions are involved in $P(n,k)$, the highest drop in the MSE is expected when many dimensions of $P(n,k)$ are activated, as it can be seen at $k=3$ ($P(1;3;4;6,3)$).

**Prototype-Based Sampling**

At a second sampling of the dynamics, with lower resolution, the materialization of the prototypes per optimization version were also analyzed. Figure 6 depicts the sequence of the produced prototypes for $k=1,2,\ldots,5$, reflecting the specific dimensions that were examined at each $k$, visualized at the scene level and the corresponding Scratch code; the stimulus from the exercise textbook used as $P_M(n)$ is also presented. From Figure 6, at $k=1$, there is a very simple output at the scene level (one sprite); however, there is more extended structure at the programing level. This reflects the activation of the GTPACK as a means for extending the gaming programing background. At $k=2$, the work at the programing level is simplified and the scene is extended with the inclusion of the second sprite. The complexity of the programing is increased at $k=3$, because there is the addition of the third sprite and the light ray. This shows higher convergence to the initial idea, revealing that the PT follows an optimization process. In addition, at $k=4$, the experimentation with the light ray and the rotation of the sources (already seen at $k=3$), along with the addition of the user feedback (score) reveal an activation of the GTPACK (to enhance the game programing skills) and of the third-person view (also seen in Figure 5). In this way, the PT sees his prototype from the eyes of the user and tries to accommodate the issues of trivial solutions and lack of feedback, thus increasing the game options and esthetics. The interpersonal interaction about esthetics continues at $k=5$, in which the scene becomes 3D, providing more immersive impression to the user. Here, it seems that the dynamics between first- and third-person views are the ones that dominate (also seen in Figure 5), as the enhancement of the external (user’s) view of the game is the focus at this stage of prototyping.

**Discussion**

**Principal Findings**

The proposed 2D-ME conceptual framework surfaces the roles of the first- and third-person views that challenge the designer to realize aspects of his/her self in this iterative procedure. Being more reflective, the designer may realize that the self may be both a subject influenced by the society norms, for example, in the aforementioned case, the programing rules, but also a driving self that may influence the society through the output of their game design. These roles were reflected in the case study presented, which served as a showcase of the 2D-ME framework to reveal the dynamics within the iterative process of the prototype optimization, seen both from the designer’s space (diary) and the game design outcome (prototype). This may contribute to more knowledgeable game designers, as it provokes reflection in a fine-grained way of thinking at different levels. It should be noted that the case study presented did not entail the hierarchy of the AT framework and was kept mostly at the activity system level, focusing on the optimization procedure.

In the case study presented, the user is a PT, who had already acquired a basic knowledge of the GTPACK framework because he works in the education sector. Nevertheless, there is no requirement for someone to have any specific level of prior knowledge to use the 2D-ME framework; simply, someone who undertakes a prototype design process (in the specific case of game design) will most probably have a basic level of the TPACK in the field. Even if he/she is not an expert, during the prototyping process, he/she will interact with the knowledge base and would acquire extra knowledge in the area. This is schematically shown in both Figure 1C and Figure 5 (GTPACK), where the context of the GTPACK is not split in equal sections but there are different levels of knowledge within it (Figure 3, c), which are dynamically expanded during the interaction with the knowledge base (Figure 5; increased GTPACK for $k=3\to4$). In addition, the knowledge of the iAIS is not provided as a constrain but as a guide to help the user formulate the internal reflection on the activities during the prototyping; someone could follow their own pathway of metacognitive process stimulation. However, we believe that the proposed iAIS targets all the reflective aspects that could be helpful to serve as a tool of memoir and reflexivity during prototyping.

Apart from focusing on the monitoring of the internal dynamics during the prototype optimization process, 2D-ME can also be used as a means of structure for the designer’s skills development in the reflection and creativity. Diary writing itself requires the designer to rethink events and processes, along with thoughts, contradictions, decisions, outcomes, which have all taken place at the various stages of the prototyping process, and “relive” it, as many times they want, even after its end. This strategy promotes their critical thinking and reasoning skills, that is, their higher-order thinking skills [39]. In addition, the stimulus from the iAIS (Textbox 1) sets an organized pathway of reflection steps that span across the triangle constructs involved in the 2D-ME framework. In this way, the activity dynamics are scaffolded by the iAIS stimulus and gradually become more imprinted in the game designer’s reflection strategy and higher-order thinking. When the narrative information of the diary is transformed into dynamics within the 2D-ME framework (Figure 5), a more externalized representation is achieved that could be visible, apart from the designer themselves, to others, as well. This can provide information for some characteristics of the designer’s personal way of thinking and creating, increasing their explainability, especially when they are not tangible at the prototype level and/or cannot be easily inferred from the final output (eg, the $P(n,M)$). In this way, the 2D-ME framework could provide more objectified basis for the analysis and comprehension of the expert designer’s process, from idea conception to the final output, which can be used for teaching the novice designers, help them to better understand their internal processes, and improve their designing skills and outcomes.

In the case study that was presented here, the prototyping process was optimized by taking into consideration the relation
between 2 adjacent versions of the prototype, that is, for the creation of \( P(n,k) \), \( P(n,k-1) \), \((k>1)\) was considered as input to the AT triangle (Figure 5). Nevertheless, additional views could also be adopted, and a higher (>1) order of memory could be used, for example, for the creation of \( P(n,k), P(n,k-3), P(n,k-2) \), \( P(n,k-1) \), \((k>3)\) can be used as a sequence of inputs. Moreover, this sequence could also vary across the whole process, both in the number of the previous versions used and in their continuity, for example, \( P(n,k-3), P(n,k-1) \), \((k>3)\). This depends on the scale of the focus that is used at each \( k \), providing the opportunity to the designer to shift from the micro- to the macroscale of the prototyping process [40,41]. This influence of previous prototype versions within the creation of the new one materializes the convolution of the dynamics adopted in the past with the dynamics shaped in the current version, in an effort to create a new, more enhanced prototype. This blending of dynamics in a macrostructural approach showcases the causalities that are created across the prototyping process that, in a way, contribute to the personal characteristics of the designer’s style. For example, sequential use of the prototyping versions (eg, as in Figures 5 and 6) could reveal steadier design style, in which each step is cumulatively constructed from the previous one. A more discontinued use of past prototyping versions could express more diffusion of the design ideas at various time instances of \( k-l (l = 1,2,... \mid k-l >0) \), revealing a distributed way of materializing the design ideas, in which different parts of them gradually are synthesized and construct a more integrated prototype version at the \( k \)th time instance.

The adopted bilateral sampling of the dynamics in the 2D-ME framework, that is, via the analysis of the diary notes and the prototype versions, provides both the internal and external views of the prototyping dynamics and outcome, respectively. Additional means of dynamics sampling could also be foreseen by incorporating other sources of relevant information, such as biosignals. In fact, the biological-signal data from multiple biological-signal sources, for example, electroencephalography, electrocardiography, electromyography, electrooculography, and electrodermal activity, can be combined with artificial intelligence or machine learning or deep learning to provide additional queues of the developer’s state (eg, stress, relaxation, or hyper or hypoactivity) during the design of the game and the creation of a series of prototypes [42-44]. In this way, the projection of the creative thinking onto the physiological signals and testing of the same protocol to game users could reveal the transferenceability of the designer’s intentions to the gamers via the similarity of the related acquired signals. This could be used as a measure to optimize the game design and create a dynamic coconstruction process [45-48], which could adapt to the specific characteristics of the designer and the target group of the serious game users [49-53].

**Limitations and Further Research**

The proposed 2D-ME conceptual framework provides new insights into the internal dynamics of the game designer’s prototyping process. Clearly, the presentation here of 1 case study with 5 prototype versions cannot express the whole range and magnitude of such dynamics. However, it serves as a good example of the potentialities the 2D-ME framework could provide, both to the game designers and educators and researchers. In addition, the adoption of MSE (equation 1), as a cost function for the optimization process (Figure 4), could also be further explored, as alternative and a combination of metrics could be involved. The selection of such cost functions should consider the acceptable variation that relates with the context of a particular applied game design problem and should clearly express, as much as possible, the gains from its minimization under the game design problem’s particular circumstances [54]. Extensive experimental application of the 2D-ME framework (eg, in educational settings of game design) could further validate its efficiency to monitor the dynamics evolved within the game design prototyping process, not only under the controlled settings, as presented here, but also at more naturalistic environments, for example, at designers’ workplaces or laboratories.

The 2D-ME conceptual framework attempts to explain the self–first and self–third person views of prototyping dynamics in serious games design. Nevertheless, as it combines the TPACK and AT frameworks in a dynamic way, it shows many potentialities to be extended further to various design problems, additionally to the area of game design. In this regard, in our future plans, we foresee 2D-ME to be applied in the areas of arts, where the game design is replaced by art design. In many areas of art, such as painting, sculpture, literature, architecture, cinema, music, and theater, the design dynamics is the driving force for the creativity and the development of the artwork through prototyping. Clearly, the 2D-ME framework can easily embrace and explain such dynamics. Furthermore, sharing of the prototype version at specific \( k \) time instances to an external group (eg, experts, colleagues, mentors, and educators) could provide an additional level of external input to the internal world of the designer, extending the role of the community. In this vein, the current 2D-ME framework can be extended to the 2D-ME-ALL one, embracing external inputs and dynamics during the prototyping process—thus, enhancing the social and collaborative aspects in the use of the AT objects and extending to runway ones with even broad societal ramifications [55]. Clearly, during the development of the prototype versions, externalization and projection from the “inner” to the “outer” world is possible and, under a cocreation perspective, wishful, to arrive at an optimized version. Moreover, this externalization does not have to be for the whole prototype but could be associated with specific parts (processes) of the prototyping, where external feedback is needed, upon the personal decision of the designer. In addition, it could happen during 1 or many iterations, providing a dynamic interplay between the inner and outer worlds. Approaching this as interconnected AT systems (constellations) [56] mainly shifts the focus of attention to the multitude of such internal activities, as externalized via their relationships, rather than the interplay between the self–first and self–third person views within the game designer’s mind. In this sense, the selection of the scale is important, and the inclusion of interconnected AT systems could explain a concept of collective contribution (ie, transition from the person to the group and, further, to society). Apparently, this can happen in the case of complementary contributions by different experts participating in the game prototyping, for example, directors, graphic designers, composers, light experts, and character...
creators. It also can be expanded to an interdisciplinary context in which the experts come from distant backgrounds (eg, health and technology). However, even in such cases, the 2D-ME framework could be adopted by each expert to describe the steps that they undertook to arrive at their final contribution to the game design. Finally, there are some creative processes that are performed alone by nature, for example, music composition and painting. In such cases, it is important to focus on the internal dynamics that take place in the artist, as, in this way, we can identify personal traits and idioms that finally characterize the personal artistic style. In this kind of cases, the constellation of AT systems could, probably, be adopted if we want to model the audience interaction with the artistic outcome, especially in cases where there is an interactive part in it, for example, music installations or interactive spaces (eg, Xenakis polytopes). These views and extensions of 2D-ME, along with the addition of biosignals’ input by incorporating artificial intelligence, machine learning, or deep learning tools, set the forthcoming research goals; work has already embarked toward such endeavor.

Conclusions
A new conceptual framework, namely 2D-ME, that provides insights upon the dynamics that evolve during a serious game designer’s prototyping process has been presented here. Combining information from both the games space, TPACK, and AT, the 2D-ME framework provides explainable representation of, somehow, intangible processes; most of them not easily identifiable in the externalization of the prototyping process, that is, the prototype itself. Using dynamic constructs, the 2D-ME framework incorporates the alternations from first-person to third-person views and vice versa, stimulated by an organized way of internal self-interview (iAIS), providing dynamics monitoring that drive the prototype optimization process. A paradigm of the practical implementation of the 2D-ME framework in the case of a game design on light reflection exemplifies its potentialities and efficiency to capture the fine-grained dynamics involved during the production of the prototype version across the design timeline. Owing to its generic structure, the 2D-ME framework has high transferability to other areas of designing, for example, arts, and could be used both as a means of exploration of the expert designer’s creative process and training of the novice ones, providing explainable representations of the underlined higher-order thinking.

Acknowledgments
The authors would like to thank the preservice teacher for his participation in the experimental case study and data sharing. This work was supported by the Khalifa University of Science and Technology, Abu Dhabi, UAE, under award number CIRA-2020-031.

Conflicts of Interest
None declared.

References


38. Scratch. URL: https://scratch.mit.edu/ [accessed 2023-03-17]


Abbreviations

AT: activity theory
CK: content knowledge
GK: game knowledge
GPCK: game pedagogical content knowledge
GPK: game pedagogical knowledge
GTPACK: Games Technological-Pedagogical-Content Knowledge
iAIS: internal Activity Interview Script
MDA: Mechanics, Dynamics, and Aesthetics
MSE: mean square error
PCK: pedagogical content knowledge
PK: pedagogical knowledge
PT: preservice teacher
TCK: technological content knowledge
TK: technological knowledge
TPACK: Technological-Pedagogical-Content Knowledge
TPACK-G: Technological-Pedagogical-Content Knowledge–Game
TPK: technological pedagogical knowledge

©Sofia Hadjileontiadou, Sofia B Dias, Leontios Hadjileontiadis. Originally published in JMIR Serious Games (https://games.jmir.org), 24.04.2023. 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.
Role of Gaming Devices Associated With Internet Gaming Disorder in China: Cross-sectional Study

Yifan Li1*, MSc; Ying Tang1*, MSc; Shucai Huang2, PhD; Linxiang Tan3, PhD; Qiuping Huang4, PhD; Xinxin Chen1, PhD; Shuhong Lin1, PhD; Zhenjiang Liao1, PhD; Hongxian Shen1, PhD

1Department of Psychiatry, National Clinical Research Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha, China
2Department of Psychiatry, The Fourth People’s Hospital of Wuhu, Wuhu, China
3Education Center for Mental Health, Central South University, Changsha, China
4Department of Applied Psychology, School of Humanities and Management, Hunan University of Chinese Medicine, Changsha, China
*these authors contributed equally

Corresponding Author:
Hongxian Shen, PhD
Department of Psychiatry
National Clinical Research Center for Mental Disorders
The Second Xiangya Hospital of Central South University
139 Renmin Road
Changsha, 410011
China
Phone: 86 13875970393
Email: shenhx2018@csu.edu.cn

Abstract

Background: Chinese gamers use computer and mobile phone games widely. Consequently, concerns regarding the development of internet gaming disorder (IGD) in adolescents have been raised. However, only a few studies have focused on the influence of gaming devices on IGD.

Objective: This study aims to compare sociodemographic information, gaming use characteristics, personality traits, and gaming motivations between computer game users (CGUs) and mobile phone game users (MGUs), as well as identifying IGD predictors.

Methods: This was a cross-sectional study. A total of 3593 internet game players took part in an online survey, which included sociodemographic information, gaming patterns, gaming motivations, the Chinese version of the Video Game Dependency Scale, and the Chinese Big Five Personality Inventory brief version. The population was divided into 2 groups for comparison by mobile phone or computer use, and the IGD population was also compared within the 2 groups.

Results: There were significant differences between the 2 gaming device groups in the time ($t_{2994}=7.75, P<.001$) and money ($t_{2994}=5.11, P<.001$) spent on gaming and in internet game addiction scores ($t_{2994}=3.68, P<.001$). Individuals using different gaming devices had different game motivations and personality traits and preferred different genres of games. Results showed that IGD predictors were different for the 2 groups, for example, strategy (odds ratio [OR] 4.452, 95% CI 1.938-10.227; $P<.001$) and action shooter (OR 3.725, 95% CI 1.465-9.474; $P=.01$) games increased the risk for MGUs.

Conclusions: Gaming devices should be considered during early identification, such as long daily gaming time, much money spent on gaming, neuroticism, and conscientiousness. In addition, more research should be conducted on new gaming devices and IGD treatment.

(JMIR Serious Games 2023;11:e40130) doi:10.2196/40130

KEYWORDS
internet gaming disorder; gaming device; gaming motivation; personality traits; gaming use characteristics
Introduction

Over the past few decades, internet games have been firmly integrated into people’s daily routines, especially among children and adolescents. According to the China Internet Network Information Center (CNNIC) latest report, by the end of December 2021, the number of internet users in China had reached 1.032 billion, and over 517.93 million netizens play games, accounting for 52.4% of the total number of internet users [1]. It is considered that video game activities improve basic attentional functions [2]. However, its inappropriate use has been considered a serious public health issue. Playing games for long periods can increase the incidence of dry-eye disease, obesity, and mental health problems [3,4]. Previous studies have found a positive association between digital device use and poor sleep quality and psychological distress, such as depression or internet gaming disorder (IGD) [5-8].

The American Psychiatric Association [9] first included IGD in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) in 2013 as a disease that needs further exploration, and in 2019, the World Health Organization (WHO) officially included it as a psychiatric disorder in the International Classification of Diseases 11th Revision (ICD-11) [10]. The reported prevalence of IGD varies due to differences in the measurement and sampling methods. Thus, the worldwide prevalence of gaming disorder is 3.05% [11]. In China, the prevalence of problematic gaming among adolescents is 5%-17% [6]. Furthermore, many studies have investigated the risk factors for IGD, but few studies have considered the influence of gaming devices; today, a large proportion of gamers use mobile phones, which allows them to play games limitlessly, which may increase the IGD risk.

Steam, the largest digital platform for gaming on a personal computer (PC), announced that 20313451 users registered on it during the coronavirus pandemic [12]. The same year, according to Google, mobile phone gamers across 10 countries increased their duration and frequency of gaming by two-thirds [13]. According to a 2020 survey, 62.5% of underage internet users in China regularly play online games. Among them, 28.9% play computer games and 56.4% play mobile phone games [14]. Owing to advances in technology, people’s entertainment has been gradually replaced by computers and video games, and many scenarios that would not happen in the real world are implemented in the virtual world, making digital games more attractive to people [15]. Games have become an increasingly popular pastime, especially during the coronavirus pandemic [16]. In today’s developed visual media, in addition to computers, mobile phones are another popular terminal for internet games. Different devices can meet different needs [17]. Computer games often feature high-quality sound and visual effects and usually require more time to play; therefore, they have the potential to provide immersion, fulfillment, and competitiveness. In addition, mobile phone games are not only easier to access than computer games but they also increase feelings of social connection and reduce feelings of isolation and negative emotions. One study found that texting games are easily associated with social networking service applications [18]. In addition, recent advances in smartphone gaming platforms have made it easier to access a wide variety of game genres [17]. Furthermore, recently, the number of mobile phone game users (MGUs) has increased dramatically, and the mobile phone gaming market has flourished, leading to a gradual shift in the traditional gaming model from desktop PC– to mobile device–based gaming models [19]. Therefore, this study aims to explore the differences between computer game users (CGUs) and MGUs to gain a deeper understanding of IGD.

In addition, exploring internet gaming use characteristics is important to better understand IGD. Internet gaming characteristics can be used as IGD predictors, including weekday/weekend game time and money spent on games, which can provide information for clinical diagnosis and treatment [20,21]. The influence of the game genre on IGD is worth mentioning. Analysis of variance reveals that internet gaming addiction varies according to the genre [22]. A noteworthy finding was that IGD might be more prevalent in users of specific genres, especially real-time strategy and first-person shooter games [23]. Gaming motivation is another important factor for predicting IGD. Escapism motivation has demonstrated a significantly strong association with IGD [24]. Other types of motivation may also be used to predict disease [25]. Thus, the theory, assessment, and treatment of IGD can be improved by examining the association between specific types of motivation [26].

Previous studies have indicated that personality traits may also play a necessary role in the development of IGD [27]. Personality traits are characterized by high stability, reflecting thoughts, attitudes, emotions, and behavioral patterns. Despite some controversy, the 5-factor model has become a widely accepted framework of personality traits, which includes 5 dimensions: conscientiousness, extraversion, agreeableness, neuroticism, and openness [28]. Personality traits can affect an individual’s experience of games, causing players to show different degrees of pathological game use disorder [29]. Gamers with IGD show high levels of neuroticism [30]. Individuals high in neuroticism tended to experience negative emotions. The act of playing games, especially in immersive environments, provides gamers with alternative virtual reality to relieve negative moods [5]. In addition, gamers with low conscientiousness have a higher risk of developing IGD because they are less persistent in pursuing goals and focusing on responsibilities in their daily lives [31]. Individuals with IGD tend to overplay games, which can affect their relationships, work, or education. Conscientious people are less likely to allow important aspects of their lives to be negatively affected by games [32]. However, results for other personality traits have been inconsistent.

Although previous studies have reported features associated with IGD, most have focused on computer games. Given that each gaming device has unique interface characteristics, we hypothesize that there are some differences in the psychometric properties and development of IGD between CGUs and MGUs. This study aims to compare internet gaming use characteristics and personality traits between different gaming devices and to explore potential specific predictors for IGD according to different gaming devices (mobile phone or computer).
Methods

Study Design and Subjects
This was a cross-sectional study that collected the internet addiction scores, game time, money, game type, game motivation, and personality traits of different gaming device users through an online survey. Differences in these indicators across users of gaming devices or gaming addicts were compared to infer the characteristics of these different populations playing games and the association between devices and gaming addiction.

This study was conducted between October and November 2019 in China. We issued online questionnaires through WeChat (a cross-platform instant messaging tool), based on a network survey. For better understanding, we used the Chinese version of the scale and the Chinese questionnaire. Participants could fill in the questionnaire anonymously by scanning the quick response (QR) code of the survey program Questionnaire Star (a free online questionnaire, survey, and voting tool). On the first page of the questionnaire, we briefly introduced the purpose of this study and some matters that needed attention when filling in the questionnaire. We also demonstrated the voluntary nature, anonymity, and confidentiality of the survey. Individuals could choose the “accept” or “reject” options. Participants could choose “reject” to exit the survey page, and those who chose “accept” would skip to the next page and continue to complete the questionnaire.

The participants were recruited online through WeChat, Weibo, or other social apps. In this study, the inclusion criteria were (1) aged 15-25 years, (2) played online games in the past 12 months, and (3) understood the questionnaire content and agreed to participate. The survey initially recruited 4012 participants. Those who had never played online games or only played offline games were excluded (n=103, 2.6%). Individuals who had severe physical or mental disorders (n=87, 2.2%) were also excluded. To be specific, the following individuals were excluded: (1) current or previous history of psychiatric disorders, including but not limited to schizophrenia, bipolar disorder, or depression, and (2) current or history of medical disorder, including but not limited to neurological and endocrinological disorders. Diagnoses of schizophrenia, affective disorders, and other disorders were obtained by experienced clinicians through the DSM-5 interview [9]. For example, if a subject presented with cognitive, affective, and behavioral disturbances and severe dissonance in clear consciousness (eg, paradoxical thinking or commentary hallucinations), schizophrenia was likely to be diagnosed; if a subject presented with significant and persistent high or low mood, depressive disorder or bipolar disorder was likely to be diagnosed. Other psychiatric disorders were also diagnosed according to the criteria in the DSM-5. In addition, we recommended that these excluded individuals seek professional medical advice. Questionnaires that lacked authenticity were deleted (eg, multiple replies from the same IP address; n=55, 1.4%). In addition, we removed responses in which the gaming device was neither a computer nor a mobile phone (eg, Nintendo Switch, Sony PlayStation, or virtual reality headsets) and individuals who played computer and mobile phone games evenly (n=174, 4.3%). Finally, a total of 3593 (89.6%) internet game users were included.

Ethical Considerations
The study protocol was approved by the Institutional Review Board. All participants were fully informed about the purpose of this investigation and signed an online informed consent form. The study was conducted in accordance with the Declaration of Helsinki, and it was approved by the Research Ethics Committee of Second Xiangya Hospital of Central South University (protocol code 20200004, date March 1, 2020).

Procedures and Measures
Self-administered questionnaires were used to collect sociodemographic information, gaming use characteristics, IGD, and personality traits. Before starting the study, the questionnaire was tested with a small sample of adolescents to improve readability and intelligibility, and it was finalized after the preliminary study.

Sociodemographic data included gender, age, education level, marital status (married or unmarried), and structure of the family (either a sibling or an only child).

In terms of gaming use characteristics, participants were asked which gaming device they used most often (mobile phone, computer, or others). Three groups were divided by reports from the responders based on the gaming time spent on either a computer or a mobile phone: (1) computer game: individuals who played only computer games or played more computer games than mobile phone games; (2) mobile phone game: individuals who played only mobile phone games or played more mobile phone games than computer games; (3) others: individuals who played computer and mobile phone games evenly or the gaming device was neither a computer nor a mobile phone. Participants were also asked about their motivation for playing games, the game genre (role-playing games [RPGs], strategy games, action shooter games, and brain and skill games), and the time (hours/day) and money (yuan/month) spent on gaming in the past 12 months. Specifically, gaming motivation included sensation seeking, escaping reality, coping with negative emotions, passing time, and making friends.

IGD was assessed using the Chinese version of the Video Game Dependency Scale (VGD-S). The scale is a revision of the English version developed by Rehbein et al [33,34] and covers the description of IGD in the DSM-5, published by the American Psychiatric Association (Cronbach’s α=.92). The scale developed by Rehbein et al [34] is a 4-point Likert-type scale (strongly disagree to strongly agree) consisting of 18 items, with 2 items for each criterion. Participants were asked to choose an experience during the past 12 months from the list of 18 descriptive items. At least 1 of the 2 items must be answered with “strongly agree” to meet the diagnostic criterion. Participants who met 5 or more of the 9 criteria were placed in the IGD group (Cronbach’s α=.90).

We used the brief version of the Chinese Big Five Personality Inventory to measure personality traits [35-37]. The scale consists of 8 items for each subtype, and there are 40 items in
total. In addition, it is scored on a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree) and demonstrates high construct validity and reliability (Cronbach $\alpha= .85$).

**Statistical Analysis**

First, to reduce bias and confounding variables, we used propensity score matching (PSM) to match CGUs and MGUs based on demographic information. Chi-square tests and $t$-tests (two tails) were used to compare demographic information, IGD scores, internet gaming use characteristics, and personality traits between CGUs and MGUs. Logistic regression analysis was performed to identify IGD predictors in different game device groups. A 2-sided statistical significance level was set at $P<.05$. SPSS Statistics version 26.0 (IBM Corp) was used in this study.

**Results**

**Sociodemographic Data**

A total of 3593 subjects were included in the analysis. Of these participants, 1511 (42.1%) were CGUs and 2032 (57.9%) were MGUs. Considering the huge number of differences, PSM was used to match the age, gender, education, marital status, and family structure (only child in the family or not). Next, we acquired 1:1-matched samples, and 1498 (99.1%) CGUs and 1498 (73.7%) MGUs were included. The mean age for each group was 19-20 years (CGU: mean 19.6, SD 1.7 years; MGU: mean 19.6, SD 1.7 years), and most of them were male (CGU: 1259/1498, 84.0%; MGU: 1270/1498, 84.8%). Almost all subjects were unmarried (CGU: 1447/1498, 96.6%; MGU: 1456/1498, 97.2%), and over half of them were an only child (CGU: 886/1498, 59.1%; MGU: 839/1498, 56.0%), which did not present a significant difference. The results of the $\chi^2$ test indicated a significant difference regarding the education level ($\chi^2= 38.3$, $P<.001$) of the 2 device-specific groups: the percentage of people who earned a bachelor’s degree or higher was greater in the CGU group (886/1498, 59.1%) than in the MGU group (717/1498, 47.9%); see Table 1.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>MGUs$^a$ (n=1498)</th>
<th>CGUs$^b$ (n=1498)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD); $t$ (df)/$\chi^2$ (df)$=0.16$ (2994), $P=.87$</td>
<td>19.58 (1.67)</td>
<td>19.59 (1.70)</td>
</tr>
<tr>
<td>Gender, n (%); $t$ (df)/$\chi^2$ (df)$=0.31$ (1), $P=.58$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1270 (84.8)</td>
<td>1259 (84.0)</td>
</tr>
<tr>
<td>Female</td>
<td>228 (15.2)</td>
<td>239 (16.0)</td>
</tr>
<tr>
<td>Marital status, n (%); $t$ (df)/$\chi^2$ (df)$=0.90$ (1), $P=.34$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>1456 (97.2)</td>
<td>1447 (96.6)</td>
</tr>
<tr>
<td>Married</td>
<td>42 (2.8)</td>
<td>51 (3.4)</td>
</tr>
<tr>
<td>Education, n (%); $t$ (df)/$\chi^2$ (df)$=38.32$ (1), $P&lt;.001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower than undergraduate</td>
<td>781 (52.1)</td>
<td>612 (40.9)</td>
</tr>
<tr>
<td>Undergraduate or higher</td>
<td>717 (47.9)</td>
<td>886 (59.1)</td>
</tr>
<tr>
<td>Only child, n (%); $t$ (df)/$\chi^2$ (df)$=1.14$ (1), $P=.28$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>839 (56.0)</td>
<td>868 (57.9)</td>
</tr>
<tr>
<td>No</td>
<td>659 (44.0)</td>
<td>630 (42.1)</td>
</tr>
</tbody>
</table>

$^a$MGU: mobile phone game user.

$^b$CGU: computer game user.

**Comparisons of Gaming Pattern IGD Score and Game Genre Between CGUs and MGUs**

Participants with IGD and the overall population were compared separately between the 2 groups (Table 2). For the overall population in the CGU group, the mean IGD score was 2.95 (SD 2.67) and the mean time and money spent were 2.21 (SD 1.56) hours/day and 111.31 (SD 189.80) yuan/month (mean US $15.95, SD US $27.20), respectively, while in the MGU group, the mean IGD score was 2.60 (SD 2.57) and the mean time and money spent were 1.95 (SD 1.26) hours/day and 63.00 (SD 149.05) yuan/month (mean US $9.03, SD US$21.36), respectively. IGD scores ($t_{2994}= 3.68$, $P<.001$), time ($t_{2994}= 7.75$, $P<.001$), and money ($t_{2994}= 5.11$, $P<.001$) in the MGU group were all significantly lower than those in the CGU group. The 2 most popular game genres in the CGU group were strategy (1048/1498, 70.0%) and action shooter (205/1498, 13.7%), whereas in the MGU group, the 2 most popular game genres were strategy (981/1498, 65.5%) and RPGs (200/1498, 13.4%). For subjects with IGD in the CGU group, the mean IGD score was 6.86 (SD 1.48) and the mean time and money spent were 3.08 (SD 2.00) hours/day and 228.34 (SD 238.15) yuan/month (mean US $32.73, SD US $34.13), respectively. In the MGU group, the mean IGD score was 6.74 (SD 1.46) and the mean time and money spent were 2.75 (SD 1.63) hours/day and 160.78 (SD 261.41) yuan/month (mean US $23.04, SD US $37.47), respectively. The time ($t_{681}= 2.37$, $P=.02$) and money ($t_{681}= 3.53$, $P<.001$) spent in the MGU group were significantly lower than MGUs. Considering the huge number of differences, PSM was used to match the age, gender, education, marital status, and family structure (only child in the family or not). Next, we acquired 1:1-matched samples, and 1498 (99.1%) CGUs and 1498 (73.7%) MGUs were included. The mean age for each group was 19-20 years (CGU: mean 19.6, SD 1.7 years; MGU: mean 19.6, SD 1.7 years), and most of them were male (CGU: 1259/1498, 84.0%; MGU: 1270/1498, 84.8%). Almost all subjects were unmarried (CGU: 1447/1498, 96.6%; MGU: 1456/1498, 97.2%), and over half of them were an only child (CGU: 886/1498, 59.1%; MGU: 839/1498, 56.0%), which did not present a significant difference. The results of the $\chi^2$ test indicated a significant difference regarding the education level ($\chi^2= 38.3$, $P<.001$) of the 2 device-specific groups: the percentage of people who earned a bachelor’s degree or higher was greater in the CGU group (886/1498, 59.1%) than in the MGU group (717/1498, 47.9%); see Table 1.
those in the CGU group. In addition, the 2 most popular game genres in both groups were strategy (CGU: 246/369, 6.7%; MGU: 222/314, 70.7%) and RPGs (CGU: 69/369, 18.7%; MGU: 47/314 15.0%).

Table 2. Within-group comparisons of IGD scores, gaming patterns, personality traits, gaming motivations, and game genres.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>MGU(^a)</th>
<th>CGU(^c)</th>
<th>Total statistics</th>
<th>IGD statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGD score, mean (SD)</td>
<td>2.60 (2.57)</td>
<td>6.74 (1.46)</td>
<td>2.95 (2.67)</td>
<td>6.86 (1.48)</td>
</tr>
<tr>
<td>Gaming patterns, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaming time</td>
<td>1.95 (1.26)</td>
<td>2.75 (1.63)</td>
<td>2.21 (1.56)</td>
<td>3.08 (2.00)</td>
</tr>
<tr>
<td>Money spent gaming</td>
<td>63.00 (149.05)</td>
<td>160.78 (261.41)</td>
<td>111.31 (189.80)</td>
<td>228.34 (238.15)</td>
</tr>
<tr>
<td>(yuan/US $)</td>
<td>(21.36)</td>
<td>(37.47)</td>
<td>(27.20)</td>
<td>(34.13)</td>
</tr>
<tr>
<td>Personality traits, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>25.43 (8.07)</td>
<td>29.68 (6.94)</td>
<td>25.22 (8.10)</td>
<td>29.18 (6.25)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>33.07 (6.33)</td>
<td>30.65 (5.26)</td>
<td>33.29 (6.57)</td>
<td>30.42 (5.65)</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>34.79 (5.98)</td>
<td>33.02 (5.59)</td>
<td>34.41 (6.56)</td>
<td>31.80 (5.86)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>29.65 (7.04)</td>
<td>28.53 (6.28)</td>
<td>30.02 (7.49)</td>
<td>28.91 (6.94)</td>
</tr>
<tr>
<td>Openness</td>
<td>32.96 (7.14)</td>
<td>31.90 (6.37)</td>
<td>34.01 (7.24)</td>
<td>32.45 (6.38)</td>
</tr>
<tr>
<td>Gaming motivations, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensation seeking</td>
<td>517 (34.5)</td>
<td>177 (56.4)</td>
<td>620 (41.4)</td>
<td>211 (57.2)</td>
</tr>
<tr>
<td>Escaping reality</td>
<td>163 (10.9)</td>
<td>86 (27.4)</td>
<td>221 (13.8)</td>
<td>14 (3.8)</td>
</tr>
<tr>
<td>Coping with negative emotions</td>
<td>673 (44.9)</td>
<td>181 (57.6)</td>
<td>711 (47.5)</td>
<td>212 (57.5)</td>
</tr>
<tr>
<td>Passing time</td>
<td>1008 (67.3)</td>
<td>193 (61.5)</td>
<td>854 (57.0)</td>
<td>204 (55.3)</td>
</tr>
<tr>
<td>Making friends</td>
<td>298 (19.9)</td>
<td>73 (23.2)</td>
<td>384 (25.6)</td>
<td>107 (29.0)</td>
</tr>
<tr>
<td>Gaming genres, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPGs(^d)</td>
<td>200 (13.4)</td>
<td>47 (15.0)</td>
<td>179 (11.9)</td>
<td>69 (18.7)</td>
</tr>
<tr>
<td>Strategy</td>
<td>981 (65.5)</td>
<td>222 (70.7)</td>
<td>1048 (70.0)</td>
<td>246 (66.7)</td>
</tr>
<tr>
<td>Brian and skill</td>
<td>149 (9.9)</td>
<td>8 (2.5)</td>
<td>66 (4.4)</td>
<td>7 (1.8)</td>
</tr>
<tr>
<td>Action shooter</td>
<td>168 (11.2)</td>
<td>37 (11.8)</td>
<td>205 (13.7)</td>
<td>47 (12.7)</td>
</tr>
</tbody>
</table>

\(^a\)IGD: internet gaming disorder.
\(^b\)MGU: mobile phone game user.
\(^c\)CGU: computer game user.
\(^d\)RPG: role-playing game.

Comparisons of Personality Traits and Gaming Motivations Between CGUs and MGUs

Table 2 outlines the significant differences between device-specific groups regarding gaming motivations and personality traits. For the overall population in both groups, CGU participants presented higher openness than MGU participants (t\(_{2994}=3.98, P<.001\)), and for subjects with IGD, CGU participants presented lower agreeableness than MGU participants (t\(_{681}=-2.76, P=.006\)). Referring to gaming motivations, including sensation seeking, escaping reality, coping with negative emotions, passing time, and making friends, the two most important game motivations were passing time (CGU: 854/1498, 57.0%; MGU:1008/1498, 61.5%) and coping with negative emotions (181/314, 57.2%), while in the MGU group, those were passing time (193/314, 61.5%) and coping with negative emotions (181/314, 57.6%).

Identifying IGD Predictors by Game Device Group

After controlling for sociodemographic factors, multivariate logistic regression was conducted to identify IGD predictors in the different game device groups (see Table 3). In the CGU group, a higher daily gaming time (odds ratio [OR] 1.301, 95% CI 1.182-1.432; P<.001), more money spent monthly on gaming (OR 1.002, 95% CI 1.001-1.003; P<.001), higher neuroticism (OR 1.060, 95% CI 1.038-1.083; P<.001), and lower
conscientiousness (OR 0.922, 95% CI 0.894-0.951; \( P < .001 \)) were associated with IGD, in addition to motivation for sensation seeking (OR 2.387, 95% CI 1.781-3.200; \( P < .001 \)), escaping reality (OR 3.407, 95% CI 2.347-4.945; \( P < .001 \)), and coping with negative emotions (OR 1.456, 95% CI 1.085-1.955; \( P = .01 \)). In the MGU group, a higher daily gaming time (OR 1.377, 95% CI 1.226-1.547; \( P < .001 \)), more money spent monthly on gaming (OR 1.004, 95% CI 1.003-1.006; \( P < .001 \)), higher neuroticism (OR 1.087, 95% CI 1.063-1.112; \( P < .001 \)), and lower conscientiousness (OR 0.928, 95% CI 0.898-0.960; \( P < .001 \)) were associated with IGD, in addition to motivation for sensation seeking (OR 2.244, 95% CI 1.651-3.050; \( P < .001 \)), escaping reality (OR 2.271, 95% CI 1.496-3.447; \( P < .001 \)), and coping with negative emotions (OR 1.493, 95% CI 1.099-2.029; \( P = .01 \)). Compared to brain and skill games, strategy (OR 4.452, 95% CI 1.938-10.227; \( P < .001 \)) and action shooter games (OR 3.725, 95% CI 1.465-9.474; \( P = .01 \)) increased the risk for the occurrence of IGD in this group, which was different from the CGU group.

Table 3. After adjusting for sociodemographic data\(^a\), binary logistic regression analyses of factors predicting IGD\(^b\) in device-specific groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CGUs(^c)</th>
<th>( P ) value</th>
<th>MGUs(^d)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gaming patterns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaming time</td>
<td>1.301 (1.182-1.432)</td>
<td>&lt;.001</td>
<td>1.377 (1.226-1.547)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Money spent</td>
<td>1.002 (1.001-1.003)</td>
<td>&lt;.001</td>
<td>1.004 (1.003-1.006)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Gaming motivations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensation seeking</td>
<td>2.387 (1.781-3.200)</td>
<td>&lt;.001</td>
<td>2.244 (1.651-3.050)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Escaping reality</td>
<td>3.407 (2.347-4.945)</td>
<td>&lt;.001</td>
<td>2.271 (1.496-3.447)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Coping with negative emotions</td>
<td>1.456 (1.085-1.955)</td>
<td>.01</td>
<td>1.493 (1.099-2.029)</td>
<td>.01</td>
</tr>
<tr>
<td>Making friends</td>
<td>1.026 (0.742-1.419)</td>
<td>.88</td>
<td>0.891 (0.614-1.294)</td>
<td>.54</td>
</tr>
<tr>
<td><strong>Personality traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>1.060 (1.038-1.083)</td>
<td>&lt;.001</td>
<td>1.087 (1.063-1.112)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.922 (0.894-0.951)</td>
<td>&lt;.001</td>
<td>0.928 (0.898-0.960)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Openness</td>
<td>0.983 (0.958-1.009)</td>
<td>.19</td>
<td>1.003 (0.971-1.031)</td>
<td>.98</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.990 (0.910-1.021)</td>
<td>.53</td>
<td>0.991 (0.959-1.025)</td>
<td>.60</td>
</tr>
<tr>
<td>Extraversion</td>
<td>1.003 (0.980-1.027)</td>
<td>.78</td>
<td>1.001 (0.973-1.031)</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Game genres</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brain and skill</td>
<td>N/A(^f)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Strategy</td>
<td>1.486 (0.612-3.607)</td>
<td>.38</td>
<td>4.452 (1.938-10.227)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Action shooter</td>
<td>1.500 (0.577-3.901)</td>
<td>.40</td>
<td>3.725 (1.465-9.474)</td>
<td>.01</td>
</tr>
<tr>
<td>RPGs(^g)</td>
<td>1.557 (0.583-4.159)</td>
<td>.38</td>
<td>2.434 (0.963-6.150)</td>
<td>.06</td>
</tr>
</tbody>
</table>

\(^a\)Sociodemographic data included gender, age, education, marital status, and family structure (only child or note).

\(^b\)IGD: internet gaming disorder.

\(^c\)CGU: computer game user.

\(^d\)MGU: mobile phone game user.

\(^e\)OR=odds ratio.

\(^f\)N/A: not applicable.

\(^g\)RPG: role-playing game.

**Discussion**

**Principal Findings**

To the best of our knowledge, this is the first study to compare internet gaming usage characteristics and personality traits among people using different gaming devices. The main research findings are as follows: First, the game time, money spent, and IGD scores in the CGU group were higher than those in the MGU group. Second, subjects in different groups preferred different game genres and had different game motivations and personality traits; for example, subjects in the CGU group performed lower openness than those in the MGU group. Third, the time and money spent on games, neuroticism, conscientiousness, motivation to seek sensations, escaping reality, and coping with negative feelings were associated with...
the occurrence and development of IGD in both groups in multivariate logistic regression; for MGUs, comparing the brain and skill, strategy, and action shooter games may bring more risks for IGD.

Several researchers have reported that time and money are dominant factors for IGD, and our study found that these 2 factors take effect in each group [20,33,38,39]. In this way, it seems that some developers of highly profitable games become whipping boys. However, in fact, most developers are just enthusiasts who want to create the most enjoyable gaming experience for their players, while game publishers are more concerned about monetization and drive developers in this way, due to which the game comes with paid content to attract players to invest as much time and money in the game as possible. Moreover, to maximize profits, the classical gaming operation pattern usually does not have fixed endpoints, which can easily cause gaming addiction [40]. The playing time, game currency input, and addiction value of the CGU group were higher than those of the MGU group. Some possible explanations are that mobile phones are currently more adapted to some medium-quality or online games with social attributes, while some computer games have higher quality, more immersive experiences, and higher prices, and high-quality games cause greater addiction problems. However, furthermore, as technology advances, the technological differences between mobile phones or other mobile devices and computers will become increasingly smaller. We can assume that mobile phone game addiction may replace computer game addiction as a larger social health issue within a few years because mobile devices are more accessible and often free to play.

The game genre plays an important role in the development of IGD, and it is important to explore its impact on IGD [41]. Our study showed that the 2 groups had different gaming preferences. For the addicted population, both groups preferred strategy and adventure games, and for MGUs, the strategy and action shooter games may bring more risks. Previous research has found that strategy games and action shooter games are popular among teenagers with high IGD risks [42]. Action shooter games have been reported to be associated with high impulsivity, disinhibition, and inattention [42]. Therefore, teenagers’ liking strategies or action shooter games need the necessary acquaintances and encouragement to engage in healthy behaviors.

Although many studies have reported the relationship between RPGs and addiction, this study did not find meaningful differences, possibly due to the heterogeneity of the subjects. The reason for the different game preferences between the 2 groups may be that the number of popular games varies on different devices, particularly when the Chinese government tightens restrictions on game distribution, which requires further research.

Certain personality traits can make individuals more prone to addiction. This study found that neuroticism and conscientiousness were both identifying factors for IGD in each group, which is consistent with previous findings [33,43,44]. Highly neurotic individuals are more likely to feel stimuli from the external environment. Online gaming is often one of the choices to gain a sense of security [31]. In addition, it is difficult for those with low levels of conscientiousness to assume appropriate responsibilities or obligations, and the online world offers an environment of less oversight and responsibility, which is tempting for them [33]. Because the 2 personality traits are significantly associated with several addiction disorders, parents and educators should not only value highly neurotic and conscientious adolescents but also help them overcome personality disorders [45,46].

We also found that the openness of the MGU group was higher than that of the CGU group. This means that people in the MGU group preferred to maintain their own gaming habits and did not easily change their game types and modes. In the IGD subgroup, the agreeableness of the CGU group was lower than that of the MGU group. Agreeableness is an assessment of the degree to which an individual likes to appear with others, and it reflects their attitude toward others. Low agreeableness can display a hostile, cynical, and ruthless attitude [47]. The finding indicates that computer game players prefer to enjoy the game alone and may encounter more social problems.

Motivations to seek sensations, escape reality, and cope with negative feelings were seen as factors strongly related to game addiction in our research, which is consistent with previous research findings [24]. In the total population, 34.5% of MGU subjects and 41.4% of CGU subjects played games because of sensation seeking, while in the IGD population, the number of people was >50%. People get pleasure from completing objectives, and developers have designed many large and small objectives in their games (eg, beating an opponent or just checking in online). Some of these in-game objectives are easy to complete, while others seem hardcore (especially those in competitive online games that require expert skills, coping, and long-term patience). These different difficulty levels allow all types of people to find their own goals in the game and earn pleasure by completing them. Some research has found that dealing with negative emotions, such as fear of failure, by imitating game characters against the real world under performance can be seen as a negative reinforcement for gaming disorder [48].

The CGU group was more motivated to look for sensations, possibly because computer games are more immersive. Players interact more with the digital world and have more fun after completing their objectives. There were >40% MGUs and CGUs whose gaming motivation was discomfort, similar to seeking sensations, among the IGD population; the 2 groups also exceeded 50%. As mentioned before, the online world offers an environment with less oversight and accountability, where players can vent their emotions unscrupulously and take less responsibility; in addition, it is noteworthy that negative emotions are often associated with high neuroticism and low conscientiousness. More than half of the participants played games to pass the time; however, interestingly, passing time did not become an IGD predictor, suggesting that most players use games in a healthy way and as a form of entertainment, and only a minority will develop IGD.

In addition, some users play games to make friends or escape reality. Some games, such as RPGs, provide a complete set of

https://games.jmir.org/2023/1/e40130
worldview settings in which people can temporarily escape from reality and meet the psychological needs of making friends [49]. This study found that people in the CGU group had a stronger motivation to make friends than those in the MGU group, reflecting that computer game players are more enthusiastic about making friends. Some studies have found that fixed social needs aggravate the progression of internet addiction [50,51]. Therefore, for computer game players, they should not only be encouraged to engage in more offline social interactions to meet their needs but also be helped in social skills.

Limitations
This study had several limitations. First, this was a cross-sectional study with a nonrepresentative sample, causality and generalizability were limited. Therefore, longitudinal research and recruiting more representative samples are necessary. Second, the research data were self-reported and came from an online survey, which may be biased in terms of the reliability and applicability of the conclusion; thus, comments about the game usage of participants’ relatives or acquaintances and offline assessment of IGD by clinicians can be included to obtain more accurate and realistic results in subsequent research. Third, considering the restrictions on game introduction in the Chinese Mainland, the study did not include people using new devices, such as PlayStation, Switch, or virtual reality headsets, and research on new devices can be designed for further study. Fourth, currently, there is no internationally recognized scale for evaluating game addiction; therefore, the results of this study should be based on preliminary results. Fifth, the study did not consider confounding factors, such as smoking, alcohol consumption, household income, and social status, which play important roles in gaming addiction and motivation. These data can be obtained in follow-up studies.

Conclusion
In summary, our study showed that there are significant differences between CGUs and MGUs in gaming usage patterns, game genres, motivations, personality traits, and internet game addiction scores. Furthermore, we found that time and money spent on games, neuroticism, conscientiousness, motivation to seek sensations, escaping reality, and coping with negative feelings are associated with the occurrence and development of IGD. Given the difference between CGUs and MGUs, there are 2 important implications for clinical research on gaming addiction disorders. More attention should be paid to high-quality computer games because of their high picture quality, immersive experience, and potential adverse consequences. The other is that useful regulation for computer game players with impaired social functioning or poor personality traits ought to be implemented. However, targeted measures need to be adopted to help people with IGD, and further studies may consider longitudinal research, including new game devices, to determine the causal relationship between addiction and various devices and other factors and accompany the innovation of game devices.

Acknowledgments
The authors would like to thank all the study participants. This work was supported by the National Natural Science Foundation of China (no. 81971249), the National Key R&D Program of China (no. 2020YFC2005300), and the Natural Science Foundation of Hunan Province (no. 2020JJ4782).

Data Availability
The data sets generated and analyzed in this study are available from the corresponding author on reasonable request.

Authors’ Contributions
All the authors contributed to the survey design. ZL and HS conceptualized and designed the research. YL and YT wrote the first draft of the manuscript and contributed to the final manuscript. SH, LT, QH, and XC prepared the assessment tools. ZL, YL, and YT performed data collection. SL and JH undertook the statistical analysis.

Conflicts of Interest
None declared.

References


47. Walia B, Kim J, Ijere I, Sanders S. Video game addictive symptom level, use intensity, and hedonic experience: cross-sectional questionnaire study. JMIR Serious Games 2022 Jun 09;10(2):e33661 [FREE Full text] [doi: 10.2196/33661] [Medline: 35471992]
Abbreviations

CGU: computer game user
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
IGD: internet gaming disorder
MGU: mobile phone game user
OR: odds ratio
PC: personal computer
PSM: propensity score matching
RPG: role-playing game

© Yifan Li, Ying Tang, Shucai Huang, Linxiang Tan, Qiuping Huang, Xinxin Chen, Shuhong Lin, Jingyue Hao, Zhenjiang Liao, Hongxian Shen. Originally published in JMIR Serious Games (https://games.jmir.org), 16.01.2023. 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.
Review

Mixed Reality in Modern Surgical and Interventional Practice: Narrative Review of the Literature

Mats T Vervoorn1*, MD; Maaike Wulfse1*, BSc; Tristan P C Van Doormaal1,2, MD, PhD; Jelle P Ruurda1, MD, PhD; Niels P Van der Kaaij1, MD, PhD; Linda M De Heer1, MD, PhD

1University Medical Center Utrecht, Utrecht, Netherlands
2University Hospital Zurich, Zurich, Switzerland
*these authors contributed equally

Corresponding Author:
Mats T Vervoorn, MD
University Medical Center Utrecht
Heidelberglaan 100
Utrecht, 3508 GA
Netherlands
Phone: 31 88 7556179
Email: m.t.vervoorn-4@umcutrecht.nl

Abstract

Background: Mixed reality (MR) and its potential applications have gained increasing interest within the medical community over the recent years. The ability to integrate virtual objects into a real-world environment within a single video-see-through display is a topic that sparks imagination. Given these characteristics, MR could facilitate preoperative and preinterventional planning, provide intraoperative and intrainterventional guidance, and aid in education and training, thereby improving the skills and merits of surgeons and residents alike.

Objective: In this narrative review, we provide a broad overview of the different applications of MR within the entire spectrum of surgical and interventional practice and elucidate on potential future directions.

Methods: A targeted literature search within the PubMed, Embase, and Cochrane databases was performed regarding the application of MR within surgical and interventional practice. Studies were included if they met the criteria for technological readiness level 5, and as such, had to be validated in a relevant environment.

Results: A total of 57 studies were included and divided into studies regarding preoperative and interventional planning, intraoperative and interventional guidance, as well as training and education.

Conclusions: The overall experience with MR is positive. The main benefits of MR seem to be related to improved efficiency. Limitations primarily seem to be related to constraints associated with head-mounted display. Future directions should be aimed at improving head-mounted display technology as well as incorporation of MR within surgical microscopes, robots, and design of trials to prove superiority.

(JMIR Serious Games 2023;11:e41297) doi:10.2196/41297

KEYWORDS
mixed reality; extended reality; surgery; intervention; education

Introduction

Over the recent years, mixed reality (MR) has gained interest within the medical community [1,2]. According to the landmark paper by Milgram et al [3], MR can best be viewed as a real-world environment enriched by virtual data presented within a single display, allowing for interaction through various means. Currently, MR is primarily experienced through head-mounted displays (HMDs) that allow virtual objects to be overlaid onto the real world using optical and video-see-through display techniques [4]. MR should be distinguished from virtual reality, referring to a completely virtual environment experienced through an immersive headset and augmented reality, which refers to virtual objects being overlaid onto a real-world environment without the possibility of interaction. More specifically, augmented reality merely projects visual data onto the real world, whereas MR anchors the visual data into the real world.
world independent of the user’s movement and allows for real-time interaction (Figure 1).

Theoretically, MR can offer many advantages for application during medical procedures or interventions by allowing the integration of relevant patient-specific data within real-time, real-world observations in a single display. It provides the ability to create interactive interfaces that facilitate procedural planning and intra-procedural navigation and can support training and education. The goal of this narrative review is to provide a qualitative overview of the different applications of MR across surgical and interventional medicine, identify its advantages and disadvantages, and reflect on its potential for future applications.

**Figure 1.** Graphical depiction of the differences between virtual, augmented, and mixed reality.

**Methods**

A targeted literature search within the PubMed, Embase, and Cochrane databases was performed on April 2, 2022, regarding the application of MR within surgical and interventional practice. Keywords in our title and abstract search included all commercially available devices for MR (Multimedia Appendix 1). Papers were screened for relevance and originality by 3 independent researchers (MTV, MW, and LMDH). Studies were included if they met the criteria for technological readiness level 5 or higher, and as such, had to be validated in a relevant environment. In short, technology readiness level is a method developed by NASA (National Aeronautics and Space Administration) in 1974 and adopted by the European Union and is used to assess technological maturation according to 9 levels, with 9 indicating the most mature level.

Papers were excluded if they involved augmented or virtual reality, for which the ability of interaction was used to differentiate between augmented reality and MR. Non-English literature and abstracts or conference proceedings were also excluded. Reference lists of all papers were screened for additional relevant literature. The flowchart of the search is shown in Figure 2. We categorized results into “preoperative and interventional planning,” “intraoperative and interventional guidance,” and “surgical and interventional training and education.”
Results

Preoperative and Interventional Planning

Within cardiovascular surgery and interventional cardiology, MR has been evaluated primarily for congenital defects that benefit from improved visualization of structural anomalies. Two studies involving congenital cardiac surgery reported a significant reduction in time required for surgical planning and intraoperative preparation when planning was done using MR instead of two-dimensional imaging [5,6]. Additionally, no intraoperative modifications to predefined surgical plans were reported in the MR group, whereas this was noted in 17.6% of cases in a 2D control group. This difference was attributed to improved spatial representation and visualization of relevant anatomy, improved depth perception, and a satisfactory correspondence to intraoperative findings in the MR group. Additionally, faster intraoperative recognition of structures was reported, presumed to be the result of improved processing of pathological structures and surgical steps by repeated visual representation beforehand [5,6]. Moreover, workload associated with mental transformation of images was reduced due to the more realistic surface features compared to 3D-printed heart models [5]. A reported future step for MR within preoperative planning of cardiac surgery would be the accurate visualization of coronary arteries and intracardiac structures and incorporating simulated movement and blood flow dynamics into the hologram.

Within orthopedic surgery, Lu et al [7] describe a collection of cases in which MR primarily improved preoperative doctor-patient communication and patients’ understanding of complex pathology.

In otolaryngology and maxillofacial surgery, studies have demonstrated improved understanding of tumor-related anatomy and planning of surgical approach when MR was used instead of conventional radiological imaging [8,9]. Mitani et al [8] found that MR facilitated recognition of dissection lines in parotid tumor surgery, where consideration of detailed maxillary anatomy is of critical importance.

Within oncological surgery, MR has been applied for preoperative planning of minimally invasive electroporation and microwave ablation for advanced gastrointestinal tumors of hepatic origin. It improved remote and hospital-based analysis of patient-specific anatomy and optimized surgical approach [10]. MR-guided surgical planning for liver resection refined understanding of hepatic vascular anatomy and tumor location, improving accuracy of resection while preserving a larger residual liver volume [11].

In urology, MR-guided preoperative planning for laparoscopic nephron-sparing resection of complex renal tumors resulted in reduced operating time and a lower conversion rate to total nephrectomy when compared to conventional 2D computed tomographic (CT)–guided planning [12,13]. The reduced operating time presumably resulted from the ability to perform a more comprehensive analysis of the renal tumor before surgery, allowing significantly more patients in the MR group
to be scheduled for laparoscopic partial nephrectomy instead of total nephrectomy (82% vs 46%; P < .001). This suggests that MR improves assessment of surgical risks and allows for modification of surgical strategy as needed, presumably by facilitating a more intuitive, stereoscopic, and comprehensive anatomical study. They also reported significantly increased patient satisfaction when MR was used during preoperative counseling [13].

In studies addressing preoperative planning, no limitations specifically related to the use of HMD are reported.

**Intraoperative and Interventional Guidance**

In neurosurgery, the efficacy of MR was reported in multiple studies that demonstrated adequate technical feasibility and safety for intracranial tumor surgery, epilepsy surgery, and spinal surgery [14-20]. Benefits mentioned include the following: (1) the ability to display holographic images of (tumor-related) anatomy onto a patient; (2) enhanced ergonomics; (3) improved preservation of attention and focus of the surgical team; (4) an intuitive workflow supported by voice commands and hand gestures; (5) cost efficiency of MR with HMD compared to conventional systems for neuronavigation; and (6) the possibility to share the surgeon’s real-time perspective with other team members and residents during surgery [14-16,18]. Overall, the reported accuracy of lesion localization was satisfactory in both intracranial and spinal procedures [16,17,20,21]. Furthermore, MR guidance demonstrated improved efficacy during transfemoral percutaneous endoscopic lumbar discectomy and external ventricular drain placement compared to conventional methods of neuronavigation, resulting in a significantly reduced operating time and exposure to radiation, increased accuracy, and a reduction of attempts required for correct drain positioning, with comparable postoperative outcome [22-24]. Interestingly, this effect was particularly strong in novice residents, practically eliminating the learning curve associated with external ventricular drain placement by performing as well as experienced surgeons while guided by MR [24]. In addition, MR-based neuronavigation for surgical treatment of hypertensive intracerebral hemorrhage has deemed feasible [25].

Several feasibility studies and case reports within interventional cardiology and cardiovascular surgery have been published that highlight positive experiences with MR during implantation of a vena cava filter [26,27] and percutaneous interventions on noncongenital pulmonary artery stenosis [28], resulting in lower doses of contrast medium and radiation exposure. This could be of particular importance in patients with reduced kidney function. A desired future development that is mentioned is the development of dynamic heart models that incorporate movement of the heart and related structures during the cardiac and respiratory cycles [29,30].

Within orthopedic surgery, MR has mostly been used in spinal procedures, such as lumbar pedicle screw implantation, operartional treatment of lumbar intervertebral disc herniation, and percutaneous kyphoplasty [31-33]. Several comparative studies report reduced bleeding and radiation exposure, shorter operating times, improved accuracy of screw placement, higher success rates, and improved pain and functionality scores when MR was compared to traditional x-ray guidance [31-33]. These results were attributed to improved anatomical understanding, resulting in smaller surgical trauma and a clearer incision site [32]. Reported disadvantages to MR include higher requirements for preoperative CT imaging, which might restrict its use to more specialized centers [31-33]. General acceptance of MR using HMD was reported as good in a study that investigated surgeon experience [34]. The devices were rated as comfortable to wear and image quality and accuracy were deemed satisfactory. Learnability of voice commands and hand gestures was well rated, although surrounding noise or lack of command understanding limited practical functionality. The potential added value for MR guidance was rated highest for surgical correction of deformities, followed by osteotomy revision and tumor surgery, especially within spinal and pelvic surgery. Greatest benefit was expected in terms of increased intraoperative accuracy, improved surgical outcome, and reduced exposure to radiation. These results demonstrate that MR can establish a sufficient support base for its application among orthopedic surgeons.

In otolaryngology and maxillofacial surgery, evidence suggests that MR enables easier extracapsular dissection of parotid tumors, which is related to lower complication rates, improved preservation of glandular tissue, and decreased incidence of saliva fistulas [9]. Tian et al [35] reported satisfactory use of intraoperative MR in cochlear implant surgery due to improved positioning of the implant. In temporal bone surgery, MR enables the surgeon to distinguish vital anatomical structures, resulting in improved surgical confidence [36]. Another reported advantage of MR is the undisrupted visual-motor axis during surgery by allowing the surgeon to preserve focus on the patient by eliminating the need for an external monitor for guidance, improving ergonomics. This was highlighted by Tang et al [37] during mandibullectomy for maxillofacial tumor resection, in which deviation from the intended surgical plane was reduced by alleviating the need to look at a monitor. They also reported improved efficiency and safety of the procedure.

Within oncological surgery, one study reported a reduction in operating time by a third and improved perceived surgical precision during minimally invasive resection of gastrointestinal tumors [10]. Moreover, MR enabled better identification of dissection lines and understanding of vascular anatomy in hepatic surgery, facilitating segmental resection [38,39]. This was highlighted in a study [11] that demonstrated MR-guided hepatectomy for hepatocellular carcinoma resulted in shorter operating time, fewer intraoperative bleeding, and a reduction in portal vein occlusion time due to an improved understanding of spatial anatomy. Furthermore, these patients had better recovery of liver function and fewer postoperative complications compared to traditional hepatectomy [11]. Percutaneous indocyanine green injection for guidance during laparoscopic anatomic liver resections was also improved when guided by MR [40]. Other feasibility studies highlight satisfactory outcomes and improved intraoperative lesion localization in breast cancer surgery [41], pediatric nephron-sparing Wilms tumor surgery [42], and robot-assisted transanal total mesorectal excision [43]. In laparoscopic cholecystectomy, MR guidance
is feasible, but it was not preferred over traditional 2D methods in one study. Especially more experienced surgeons responded neutrally or negatively toward the implementation MR for laparoscopic cholecystectomy due to increased operating time and no significant improvements in outcome [44]. Lastly, in an extensive survey-based study [45], the reported overall experience with MR using HMD was good, with an important advantage being the ability to superimpose (visual) information on the patient’s body. They concluded that MR using HMD improved the overall speed and comfort of a surgical procedure by providing intraoperative support and additional guidance to the surgical team [45]. A reported disadvantage of concern to MR is the inability to integrate the real-time movement of abdominal organs and the effects of soft-tissue deformation into the hologram, limiting translation of preoperatively acquired images to the real-time surgical site [38]. This is especially problematic during minimally invasive procedures and requires the development of complex prediction models to overcome these problems.

Evidence from comparative studies supports the use of MR within urology [12,13]. In laparoscopic nephron-sparing surgery for complex renal tumors, MR resulted in reduced operating time, warm ischemic time associated with renal artery clamping, reduced estimated blood loss, and a lower conversion rate to total nephrectomy when compared to a conventional approach with 2D CT guidance. Other factors such as inhospital stay, postoperative serum creatinine, and complication rate were similar among groups. The authors attributed the observed benefits to improved intraoperative identification of relevant anatomical structures, such as a tumor’s feeding arteries, thereby preventing clamping of the main renal artery [12,13].

Within interventional radiology, Deib et al [46] conducted a feasibility study into MR-guided percutaneous spine procedures. They concluded that MR using HMD was noninferior to guidance by traditional monitors and might prove especially valuable during procedures that are conducted in spatially limited environments, as key anatomic landmarks could be reliably projected onto the procedural site, which limits disruption of the visual-motor axis.

MR might also be beneficial in plastic and reconstructive surgery due to the ability to project holographic visualizations of anatomical structures onto the patient to guide reconstruction. An example of this was a study [47] into auricular reconstruction, in which an image of the contralateral auricle was projected onto the ipsilateral surgical site. After appropriate positioning was ensured, MR resulted in improved perceived reconstructive results compared to a conventional approach with transparent film guidance [47]. Application of MR during vascular flap transfers and reconstructions that require identification and preservation of target perforator vessels, such as deep inferior epigastric perforator flap reconstruction of breast tissue, resulted in improved localization of the target perforating vessels when compared to Doppler-ultrasound [48]. An encountered problem, however, was related to adequate depth perception of the target vessels, which was limited using MR [49].

Within pulmonary surgery, MR can be used to visualize the location of small pulmonary nodules in patients scheduled for resection by video-assisted thoracoscopic surgery based on images acquired by preoperative CT. Using HMD, one study [50] reported that 94% of nodules were accurately localized, compared to 30% by manual palpation. This is especially relevant, as failure to localize a target nodule might lead to unplanned lobectomy or sampling error. Another study [51] reported the ability to integrate simulated lung deflation into the hologram, which improved surgical instrument placement and allowed for easier identification of nonpalpable lesions, which are notoriously difficult to localize once the lung is deflated [50,51].

Although the above demonstrates the feasibility of MR using HMD, some limitations specifically related to the use of HMD are reported. A recent meta-analysis proposed that using HMD during intracranial tumor surgery might provide additional challenges regarding depth perception, possibly increasing inaccuracy in surgery with small target lesions [52]. These issues restrict HMD use during intracranial procedures that require navigation on submillimeter level and are performed with a microscope. A solution could be the integration of preoperatively acquired holograms into the visuals obtained through the surgical microscope instead of HMD [15]. A reported drawback in cardiovascular surgery is the lack of integration with other often used head-worn devices, such as surgical loupes and headlamps [29]. Other reported limitations related to HMD are as follows: (1) holographic drifts while walking around the hologram [14]; (2) delayed image tracking with fast head movement [37,53]; (3) the perceived parallax effect [45]; (4) holograms being affected by surgical light [37,53]; and blind spots caused by obstruction of the surgical field by the hologram [14,54]. The latter, however, seems manageable with adjustable hologram opacity [14]. Another technological limitation is the relatively short battery life of the HMD, which poses difficulties for long surgeries [37,45,53]. Ergonomic disadvantages were related to perceived added strain on the musculoskeletal system of head and neck, eye strain, and visual discomfort [45,54].

**Surgical and Interventional Training and Education**

Due to its inherent qualities, MR could benefit surgical and interventional training and education. Condino et al [55] developed a multimodal MR-based surgical simulator for hip arthroplasty. Authors reported that this improved perception of spatial relationships between real and virtual objects. Such a simulator could be of relevance, since hip arthroplasty accounts for a multitude of reported adverse events in orthopedics, and risk of complications during the procedure is strongly related to surgeon experience.

Within otolaryngology and maxillofacial surgery, MR facilitated surgical training of maxillary carcinoma resection, and its overall usefulness was rated 4.5 on a 5-point scale in a survey study among otolaryngologists. This was attributed to improved understanding of the surgical procedure [8]. Additionally, MR can provide a training method for transcanal endoscopic ear surgery and cochlear implantation [36].

The possibility to establish a bidirectional audiovisual connection between the observer wearing an HMD and other
participants at a remote site offers unique opportunities for training and education, as it allows medical students and residents to participate in (surgical) ward rounds and procedures remotely from a first-person perspective, providing teaching that was otherwise inaccessible to students [56]. This was recently supported in a study regarding introduction of an HMD within grand surgical rounds [57]. Moreover, it allows for telementoring through data sharing and communication with other surgical team members or surgical residents [45]. Telementoring by off-site experts using HMD has already been demonstrated in neurovascular procedures and cancer surgery [58,59].

Furthermore, evidence suggests a benefit to patient education as well, as an increase in comprehension and a decrease in anxiety was experienced by patients after being subjected to MR-based counseling [60], while simultaneously providing a potential method of intervention to improve patients’ understanding of their own disease and adherence to treatment, and hence, facilitate knowledge transfer between health care professionals and individual patients [61].

In studies addressing surgical and interventional training and education, no limitations specifically related to the use of HMD are reported.

**Discussion**

The introduction of MR seems to have impacted the role of technology within surgical and interventional practice. The increasing number of studies on MR-based modalities using HMD has accelerated this development. It is likely that this number will continue to grow in the following years, mirroring the technological developments that succeed each other in rapid pace. Based on our findings, we can hypothesize that MR-based strategies for preoperative planning and intra procedural guidance have certain benefits including, but not limited to the following: (1) challenging procedures with high anatomical complexity; (2) procedures that are performed infrequently or are otherwise highly specialized and of low volume, such as rare congenital cardiac defects; (3) procedures that involve an extensive learning curve; (4) procedures that rely on extensive preoperative imaging for intra procedural guidance; and (5) general education and skill training purposes. Additionally, an increasing number of studies are emerging that highlight the feasibility and potential superiority of MR for more common procedures. In this regard, it is important to note that the benefits of this technology, although complex in nature, do not have to be limited to complex, uncommon procedures but can also further improve safety and efficacy of more simple and routine procedures by improving workflow.

An often reported benefit of MR is related to improved efficiency, which could be the result of improved planning and execution. The immersive, high-quality holographic models of patient-specific anatomy that are compatible with the newest HMD facilitate interaction and allow for improved understanding of complex spatial relationships between relevant anatomical structures, improved spatial orientation, and identification of target pathologies, thereby enabling better visualization and preparation beforehand. As modern medicine is moving toward personalized precision treatment, these patient-specific holograms and qualities of MR could further enhance individually customized surgical plans, including decision-making regarding surgical approach. During a procedure, the ability to project a reconstructed holographic image onto a patient allows for better identification of important patient-specific anatomical landmarks, prevents disruption of the visual-motor axis, and intuitively improves the accuracy of most interventions, resulting in improved surgical performance and reduced operating time. As many procedures that rely on real-time imaging and monitoring for guidance are radiation-based, improved efficiency can result in a significant reduction in total radiation exposure to both patient and physician. This is further illustrated by an experimental model that was developed for real-time radiation exposure dose visualization [62]. Patients and health professionals regularly exposed to radiation could benefit from such a model by creating awareness and ultimately avoiding unnecessary radiation, although validation in a relevant environment is still needed. Furthermore, by decreasing the need for contrast fluids, patients with reduced renal function could benefit as well. Data sharing options could result in better cooperation between team members and improve telementoring and remote counseling by facilitating transfer of knowledge between experienced physicians and resident doctors, thereby improving skill and merit in the inexperienced colleague. Besides, MR seems to possess added value for medical education, skill training, and patient counseling, as suggested by multiple reports that support improved efficiency and engagement as well as increased surgical confidence and skill level when MR-based technologies are used. Ultimately, this could result in more skilled surgeons in the operating room, decreasing the complication risk associated with human inexperience.

Although the overall experience with MR is rated positively in most studies, drawbacks of this technology seem primarily related to HMD and include ergonomic and visual concerns through the added strain of an HMD on head and neck musculature as well as visual fatigue. The experienced field of view might be limited, and occurrence of the parallax effect has been described, which refers to the phenomenon when content in the background moves at a different speed compared to content that is positioned on the foreground [45]. For certain areas that involve dynamic organs, the static nature of holograms based on preoperatively acquired imaging is perceived as an important hurdle for satisfactory and accurate intraoperative guidance, as well as problems related to soft-tissue deformity and disturbed depth perception. These problems hinder the application of MR during intra-abdominal and cardiovascular procedures and warrant the development and integration of prediction models for soft-tissue deformation and dynamic movement of organs into the displayed holograms. In that regard, the development of technology that allows the integration of simulated lung deflation into holograms for pulmonary surgery seems promising [51]. Other reported minor disadvantages are related to battery life, delayed image tracking with swift head movements, and the need for additional training to tolerate eye strain and visual discomfort caused by the hologram. However, these disadvantages do not seem to impact the generally reported positive experience of MR technology.

https://games.jmir.org/2023/1/e41297

JMIR Serious Games 2023 | vol. 11 | e41297 | p.554

(page number not for citation purposes)
and we expect most of these problems to be solved with ongoing technological development or integration within other emerging technological supportive tools for surgery, such as the surgical robot and surgical microscopes, which seem less susceptible to most reported HMD-related drawbacks of current MR technology. In this regard, it seems of utmost importance that we do not blind ourselves for the potential applications of MR beyond HMD to further accelerate the advancement of this technology.

Since the introduction of HMD for MR, they have evolved from heavy, obstructive, and wired devices to become lighter, see-through, and wireless. Especially the emergence of the Hololens (Microsoft) has offered a more immersive experience compared to previous generations of HMD and has accelerated both innovation and interest in the usefulness of this technology. As a result, almost all studies use Hololens as the designated HMD, whereas the use of other commercially available HMD within health care is limited. Given the value of competition for technological advancement, a more heterogeneous field of HMD suppliers besides Microsoft would be desirable to improve and accelerate development of this technology. Future developments should include the integration of MR within images acquired through surgical microscopes, in robotic surgery, and in the construction of holograms based on real-time data with prediction models that reflect the dynamic nature of organs as opposed to holograms based on preoperatively acquired static imaging. This seems a prerequisite next step for the maturation and adoption of MR technology in current clinical practice and can be designated as the next frontier to overcome for this technology. Besides, focus should be aimed at designing clinical studies to validate the superiority of MR-guided procedures compared to conventional ones. This warrants specific outcome parameters that assess outcome on both the surgical (for example NASA task load index) and patient side (composite end points), which obviate the need for an extensive sample size to prove superiority and make clinical trials more executable.

Limitations to this narrative review are related to the qualitative nature of the paper, which limits its level of evidence. Therefore, the paper should primarily be viewed as a general summary of the topic and a document that could be used to guide future research. The results of our search might also be troubled by the lack of a clear and universally applied definition of MR. As we have noticed, the terms augmented reality and MR seem to be used interchangeably in the literature. We hypothesize that this finding is primarily related to the novelty of the technique, and we expect that, as development progresses, they will become more established in clinical practice.

In conclusion, the implementation of MR seems to possess certain benefits, primarily related to efficiency and accuracy by facilitating preoperative planning and intraoperative guidance, especially in complex, low-volume cases that involve complex anatomy. However, this does not preclude its use in more common, less complex procedures. Besides, it may also benefit surgical training and education of younger residents and peers. Areas of improvement seem to be primarily related to issues involving the use of HMD, which warrant attention to applications beyond HMD in future developments.

Acknowledgments

MTV and MW were responsible for the search, screening, and writing of the manuscript. TPCVD, JPR, and NPVdK provided advice and reviewed the manuscript. LMDH contributed to the search and reviewed the manuscript. All authors approved the final version before submission.

We would like to thank Floor Patist for her contribution to Figure 1.

Conflicts of Interest

TPCVD is the founder and current Chief Medical Officer of Augmedit B.V., Naarden, The Netherlands.

Multimedia Appendix 1
Keywords in title and abstract search.
[DOCX File. , 13 KB - games_v11i1e41297_app1.docx ]

References


Abbreviations

CT: computed tomography
HMD: head-mounted display
MR: mixed reality
NASA: National Aeronautics and Space Administration

https://games.jmir.org/2023/1/e41297
Review

Children and Young People’s Involvement in Designing Applied Games: Scoping Review

Michael John Saiger¹, BA, MPhil; Sebastian Deterding², MA, PhD; Lina Gega³, RMN, BA, BN, PhD

¹Department of Computer Science, University of York, York, United Kingdom
²Dyson School of Design Engineering, Imperial College London, London, United Kingdom
³Department of Health Sciences & Hull York Medical School, University of York, York, United Kingdom

Corresponding Author:
Sebastian Deterding, MA, PhD
Dyson School of Design Engineering
Imperial College London
25 Exhibition Road
South Kensington
London, SW7 2DB
United Kingdom
Phone: 44 020 7594 8888
Email: s.deterding@imperial.ac.uk

Abstract

Background: User involvement is widely accepted as key for designing effective applied games for health. This especially holds true for children and young people as target audiences, whose abilities, needs, and preferences can diverge substantially from those of adult designers and players. Nevertheless, there is little shared knowledge about how concretely children and young people have been involved in the design of applied games, let alone consensus guidance on how to do so effectively.

Objective: The aim of this scoping review was to describe which user involvement methods have been used in the design of applied games with children and young people, how these methods were implemented, and in what roles children and young people were involved as well as what factors affected their involvement.

Methods: We conducted a systematic literature search and selection across the ACM Digital Library, IEEE Xplore, Scopus, and Web of Science databases using State of the Art through Systematic Review software for screening, selection, and data extraction. We then conducted a qualitative content analysis on the extracted data using NVivo.

Results: We retrieved 1085 records, of which 47 (4.33%) met the eligibility criteria. The chief involvement methods were participatory design (20/47, 43%) and co-design (16/47, 37%), spanning a wide range of 45 concrete activities with paper prototyping, group discussions, and playtesting being the most frequent. In only half of the studies (24/47, 51%), children and young people participated as true design partners. Our qualitative content analysis suggested 5 factors that affect their successful involvement: comprehension, cohesion, confidence, accessibility, and time constraints.

Conclusions: Co-design, participatory design, and similar high-level labels that are currently used in the field gloss over very uneven degrees of participation in design and a wide variety of implementations that greatly affect actual user involvement. This field would benefit from more careful consideration and documentation of the reason of user involvement. Future research should explore concrete activities and configurations that can address the common challenges of involving children and young people, such as comprehension, cohesion, confidence, accessibility.

[(JMIR Serious Games 2023;11:e42680)](https://games.jmir.org/2023/1/e42680) doi:10.2196/42680

KEYWORDS

serious game; game design; end user; participatory design; co-design; user involvement method; interventions; game development; children; pediatric; young people; child; youth; review method; scoping; applied game
Introduction

Applied Games

Applied games, or serious games, describe (usually digital) games designed to drive desired cognitive, behavior, or other changes in players and the wider community [1-4]. In recent years, applied games have seen increasing interest in areas, such as well-being [5], mental health [6], and education [7,8].

Applied games markedly differ from entertainment games in that they need to fit the capacities, needs, preferences, and contexts of often highly idiosyncratic target end-user audiences; find game mechanics and content that are both appealing or “fun” to the target audience; and deliver the “active ingredients” producing the desired changes, be it learning content, persuasive messages, or medical treatment regimes [4,9]. This is particularly true for applied games targeting children and young people (CYP), whose general and gaming capacities, preferences, and contexts not only drastically differ from those of adult designers and players but also from each other, depending, for example, on their developmental stage [3,10].

As in general design and development, one major successful strategy for sensitizing applied game designers to the specifics of target end users is to directly involve end users (and other relevant stakeholders) in the design, implementation, and evaluation of the game in question [11-13]. Involving users in the design process has been shown to improve use and treatment engagement [13,14], usability [15], and system adoption and adherence across stakeholders [16-18]. On the side of developers, it promises improved understanding of user needs, reduced development costs and time, and improved design quality [19]. Consequently, there are growing calls for regularly using user involvement methods in the design of applied games for health [10,11,20].

User Involvement

However, “user involvement” describes a wide and messy field. Different research and practice communities have developed parallel traditions with confusing differences and overlaps in name, underlying values, and details of implementation, for example, human-centered design [15,21] in computing and human-computer interaction (HCI); participatory design [22,23], co-design [21], or coproduction [24] in design; patient and public involvement in health [25,26]; or action research, participatory research and science, or citizen science across the (social) sciences [15,25,27-29]. Despite regularly involving end users in the form of playtesting, applied game design still has no strong tradition of granting users more agency and input, particularly in the early stages of the design process [30,31]. In addition, recent analyses have shown that digital mental health intervention projects vary greatly in how they approach user involvement and often fail to document how user involvement methods are implemented in detail [10,20]. Bergin et al [16] observed in their recent review that few studies similarly reported any user experience of the involvement processes used—there is no even consensus on how the user experience of study participation should be captured. As a result, there is presently little empirical data on how differences in the detailed implementation of user involvement would affect end-user engagement and the overall efficacy of involvement. Consequently, we do not have evidence-based guidance on how to best involve end users and other stakeholders in the design of applied games [32], for example, for mental health [1,16,33] or education [8,34].

This lack of evidence and guidance is particularly pertinent for user involvement of CYP, and we neither know what roles, activities, or tools help engage and maintain their participation [17,29] nor what implementation factors would affect their involvement or actual adoption, adherence to, and efficacy of developed interventions (an exception was found in a study by Shah et al [35]).

Research Questions

Thus, before we can begin to articulate potential best practice guidelines for the implementation of user involvement in designing applied games with CYP, we need basic stocktaking of actual existing practice and evidence. To this end, this scoping review aimed to describe how user involvement has been implemented in designing games for CYP and what factors (if any) likely affect effective CYP involvement. We articulated this aim in 4 research questions (RQs):

- RQ1: What user involvement methods are used for what purpose?
- RQ2: In what roles are CYP being involved?
- RQ3: How are user involvement methods implemented in detail?
- RQ4: What factors affect effective CYP involvement?

Methods

Study Design

For our scoping review, we combined a systematic literature review [36] with inductive qualitative content analysis [37]. We report our method following the revised 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [36]. We did not preregister this scoping review because of the descriptive and exploratory nature. All study materials can be found in the Open Science Framework repository (Multimedia Appendix 1).

Eligibility Criteria

As noted, our review focuses on user involvement methods in the design of applied games with CYP. With “user involvement methods,” we capture any approach that involves end users in the design of a game, such as participatory design, co-design, play-centric design, or user-centered design. With “applied games,” we mean any attempt to create or use game-based software for a nonentertainment change. “CYP” was interpreted verbatim from the records that captured a variety of age ranges, for example, 2 to 4 years or 16 to 25 years.

Unfortunately, there are no well-established standard terms or controlled vocabularies for our search focus; for example, applied games are commonly called “serious game,” “game-based,” “game for X,” or “gamified X.” Following similar reviews [10,11,38], we developed a core search string that combined a range of synonymous keywords for each aspect,
such as CYP, games, applied contexts, and user involvement methods (Textbox 1).

We analyzed papers published in the last 10 years (January 1, 2010, to December 7, 2021), and as the field of applied games is quite new, we did not expect a great number of texts before this date. We implicitly focused on selecting papers that had a specified “applied context” (Textbox 1) in which, for example, games were developed with young people for “mental health” or “education.” Furthermore, we focused on the last 10 years of records owing to the development and changes in technologies and methodologies. Capturing recent (within 10 years) studies provides an up-to-date overview of the state of the literature.

We included full length papers that reported on the design of game-based software targeting CYP and explicitly featured and reflected on user involvement methods. As we are interested in understanding the factors affecting user involvement, we excluded papers that did not reflect or evaluate the user involvement methods they used. We did not exclude papers based on application contexts—we wanted to avoid unnecessary focus on a specific domain, such as health-related literature, because insights may already have been made in educational gaming or entertainment game design but ignored by us because of a narrow focus on health-related literature. We excluded short or work-in-progress papers, because they did not provide enough space for detailed reflection and reporting on the methods used. We did not investigate gray literature because of the difficulty of creating a reliable and reproducible search strategy for such a dispersed and unstructured collection of materials with no core databases. See Textbox 2 for the full inclusion and exclusion criteria.
Textbox 1. Full list of used synonymous key terms for constructing search strings, in which a search string combines all terms, that is, children and young people (CYP) synonyms AND games synonyms AND applied context synonyms AND user involvement method.

<table>
<thead>
<tr>
<th>CYP</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Young people</td>
</tr>
<tr>
<td>• Young adults</td>
</tr>
<tr>
<td>• Students</td>
</tr>
<tr>
<td>• Kid</td>
</tr>
<tr>
<td>• Child</td>
</tr>
<tr>
<td>• Adolescent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Games</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Game</td>
</tr>
<tr>
<td>• Video game</td>
</tr>
<tr>
<td>• Computer game</td>
</tr>
<tr>
<td>• Gamified</td>
</tr>
<tr>
<td>• Game based</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Applied contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mental health</td>
</tr>
<tr>
<td>• Mental disorders</td>
</tr>
<tr>
<td>• Anxiety or depression</td>
</tr>
<tr>
<td>• Therapy</td>
</tr>
<tr>
<td>• Cognitive behavioural therapy</td>
</tr>
<tr>
<td>• Human computer interaction (HCI)</td>
</tr>
<tr>
<td>• Education</td>
</tr>
<tr>
<td>• Learning</td>
</tr>
<tr>
<td>• Behaviour change</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User involvement method</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Co-creation</td>
</tr>
<tr>
<td>• Co-design</td>
</tr>
<tr>
<td>• Co-production</td>
</tr>
<tr>
<td>• Participatory design</td>
</tr>
<tr>
<td>• Patient centred design</td>
</tr>
<tr>
<td>• Patient partner involvement</td>
</tr>
<tr>
<td>• User-centred design</td>
</tr>
<tr>
<td>• Person-centred</td>
</tr>
<tr>
<td>• Collaborative</td>
</tr>
<tr>
<td>• Iterative design</td>
</tr>
<tr>
<td>• Cooperative</td>
</tr>
</tbody>
</table>
**Textbox 2. Inclusion and exclusion criteria.**

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Article features user involvement methods</td>
</tr>
<tr>
<td>• Article reports on the design of game-based software</td>
</tr>
<tr>
<td>• Game-based software targets audience of children or young people</td>
</tr>
<tr>
<td>• User involvement methods are evaluated or reflected on</td>
</tr>
<tr>
<td>• Full paper</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No user involvement method</td>
</tr>
<tr>
<td>• No reflection or evaluation of user involvement method</td>
</tr>
<tr>
<td>• Not related to game-based software</td>
</tr>
<tr>
<td>• Not published in English</td>
</tr>
<tr>
<td>• Not a full or original paper (eg, work in progress, conference summary, or workshop)</td>
</tr>
<tr>
<td>• No children and young people involved</td>
</tr>
<tr>
<td>• Not retrievable</td>
</tr>
</tbody>
</table>

**Information Sources and Search Strategy**

We searched 4 databases: the ACM Digital Library, IEEE Xplore, Scopus, and Web of Science. These databases mirror the interdisciplinary structure of research on applied games: the Association for Computing Machinery Digital Library and IEEE Xplore, which cover computing and HCI conferences and journals in which the bulk of technical games research is published, while Web of Science and Scopus capture disciplines such as medical research, psychology, and education.

Before starting our search, we iterated on variations of search strings implementing our target keywords (Textbox 1) for each database, because each database afforded different search strings tools. The final search strings are presented in Multimedia Appendix 2. We conducted the first full search on May 7, 2021, and the last search on July 5, 2021.

The studies retrieved from the databases were managed using State of the Art through Systematic Review (StART; version 3.3, Beta 03; Laboratório de Pesquisa em Engenharia de Software). StART identified additional studies through snowballing, which were added to the selection for screening. In addition, 14 studies were manually added.

**Selection Process**

We first removed duplicate records using StART, which identified duplicates across databases. Additional duplicates were then manually removed from the selection. The first author then manually screened titles, keywords, and abstracts against the eligibility criteria; sourced full texts of the eligible studies; and then manually assessed full texts for eligibility. Finally, records that were reported in the same study were merged.

**Data Items and Collection Process**

All relevant information was extracted in StART, which was then exported (xlsx format) into the qualitative data analysis software NVivo (version 12; QSR International) for open coding across the extracted information. For each eligible study, the first author extracted standard metadata (title, authors, abstract, and year) in addition to a range of descriptive data.

To describe our sample and study characteristics, the first author coded papers by the following parameters:

1. **Discipline**: disciplines were coded first verbatim by the title and self-description of the publication venue and then inductively aggregated; for example, a paper published in ACM HCI was coded as “human-computer interaction,” as it describes itself as “The ACM CHI Conference on Human Factors in Computing Systems is the premier international conference of Human-Computer Interaction (HCI).”
2. **Date of publication**: extracted from paper metadata
3. **The country where the study was conducted**: extracted verbatim from the Methods section
4. **The number of CYP involved**: extracted from the Methods section of the paper
5. **The age and age range of CYP involved**: extracted from the Methods section of the paper
6. **The kind and number of participant groups involved**: participant groups (eg, children, parents, and clinicians) were first extracted verbatim from the method sections, and then the number of different groups was counted.

To describe user involvement methods and roles, the first author coded papers for the following factors:

1. **The self-labeled user involvement methods used**: extracted verbatim from how the authors labeled their study in the title or Methods section; this resulted in multiple labels for some studies in which terms were used interchangeably.
2. **The authors’ stated aims of user involvement**: first extracted verbatim from the “goal” or “aim” statements of each paper, then inductively coded into higher-level categories, such as “determine features and functionality” or “explore methodology”; 1 paper could entail multiple aims.
3. In what capacity were CYP involved: first extracted as verbatim labels given to their roles in the paper (Multimedia Appendix 2); where roles were not explicitly labeled in the paper, we noted them as “not stated.” On the basis of the full paper description of children’s involvement, we inductively mapped each study on the Druin [39] 4-fold typology of children’s roles in the development of new technologies.

Finally, to describe how user involvement was implemented and how to identify emerging factors affecting it, we imported the extracted data fields and full-text PDF documents into NVivo for inductive qualitative content analysis [37]. The first author inductively coded the method, discussion, and conclusion sections of papers for related emerging themes in the first and second focused coding cycle [40]. Descriptively, (10) study structure, (11) activities, and (12) media and tools have emerged as high-level categories. Of these, the activities were clear and distinct enough that we could conduct a follow-up frequency count. In terms of (13) factors affecting user involvement, 4 themes emerged.

**Bias and Certainty Assessment**
Because the aim of our study was narrative description, not establishing summary effects, no bias or certainty assessments were performed.

**Synthesis of Results**
For descriptive summary reporting, we calculated the frequencies for (1), (2), (4), (5), (6), (7), and (11). Emerging themes and observations were synthesized through standard inductive qualitative content analysis.

**Results**

**Study Selection**
Our search returned 1085 records. Title and abstract screening removed 81 duplicate records and further excluded 885 records, of which 164 (18.5%) did not clearly incorporate user involvement methods, 626 (70.7%) lacked explicit reflection of user involvement methods, 63 (7.1%) were not related to game-based software, and 32 (3.6%) were not full papers. At the full-text stage, of the remaining 13.4% (119/885) of records, we excluded a further 69 (58%), of which 22 (32%) were not full papers, 12 (17%) lacked user involvement methods, 22 (32%) lacked a detailed reflection of them, 4 (6%) had no CYP involvement, and 1 (1%) was not related to games; we could not obtain access for 7 (10%) records, and 1 (1%) was not published in English. Of the 50 remaining papers, we merged 3 (6%) paper pairs that reported the same study, resulting in 47 final studies in the analysis. Figure 1 presents the PRISMA flow diagram.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. CYP: children and young people.
Study Characteristics

Table 1 presents the characteristics of the included studies; 47% (22/47) of the studies were from computer science and HCI, 19% (9/47) were from education, 17% (8/47) were from serious games or games design, 11% (5/47) were from health, and 6% (3/47) were from psychology. There was no clear upward or downward trend in the publications over time. In total, 85% (40/47) of studies were conducted in the Global North, led by European countries (15/47, 32%), then the United Kingdom (13/47, 28%), Australia and New Zealand (7/47, 15%) and the United States (4/47, 9%), followed by Brazil (3/47, 6%) in the Global South.

The sample size of CYP involved ranged from 2 to 109, with an average of 23 (SD 24.8) and a median of 13. Table 2 shows the frequency of sample sizes with 26% (12/47) studies involving 6 to 10 participants, 23% (11/47) involving 11 to 25 participants, 19% (9/47) involving 26 to 50 participants, 17% (8/47) involving 1 to 5 participants, and 15% (7/47) involving >51 participants. This is likely because (1) most studies focused on interventions for intersectional groups, such as children with learning difficulties, in which sampling can be challenging; and (2) user involvement methods often value “thick data” with small-sample qualitative methods and a struggle to scale to large participant numbers. Nevertheless, many studies have reported “small sample sizes” as a limitation.
<table>
<thead>
<tr>
<th>Study</th>
<th>What method or methods were labeled or framed?</th>
<th>Participant groups, n</th>
<th>Groups involved</th>
<th>Discipline</th>
<th>How long was the study?</th>
<th>Median age (years)</th>
<th>Young people, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>All et al [41], 2012</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, teachers, other experts, designers and developers, and researchers</td>
<td>Interaction design</td>
<td>Unknown</td>
<td>17.5</td>
<td>72</td>
</tr>
<tr>
<td>Alves and Hostins [42], 2019</td>
<td>Design-based research</td>
<td>3</td>
<td>CYP, designers and developers, and researchers</td>
<td>HCI</td>
<td>4-6 months</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Al-Wabil et al [43], 2010</td>
<td>User-centered design</td>
<td>4</td>
<td>CYP and clinicians</td>
<td>HCI</td>
<td>Unknown</td>
<td>15.5</td>
<td>6</td>
</tr>
<tr>
<td>Anacleto et al [44], 2012</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, clinicians, and support group</td>
<td>HCI</td>
<td>2-3 months</td>
<td>9.5</td>
<td>2</td>
</tr>
<tr>
<td>Anthony et al [45], 2012</td>
<td>Participatory design</td>
<td>4</td>
<td>CYP, designers and developers, and researchers</td>
<td>Game design</td>
<td>&lt;1 day</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Aufegger et al [46], 2020</td>
<td>Co-design</td>
<td>4</td>
<td>CYP, carers and family members, designer and developer, clinicians, and researchers</td>
<td>Computer science</td>
<td>2-3 months</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Benton et al [47], 2012</td>
<td>Participatory design</td>
<td>5</td>
<td>CYP, teachers, and researchers</td>
<td>Health</td>
<td>2-3 months</td>
<td>12.5</td>
<td>12</td>
</tr>
<tr>
<td>Benton and Johnson [48], 2014</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, carers and family members, teachers, and researchers</td>
<td>Education</td>
<td>2-3 months</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Bonsignore et al [49], 2016</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, other experts, designers and developers, and researchers</td>
<td>Interaction design</td>
<td>10-12 months</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>Bossavit and Parsons [50], 2016</td>
<td>Participatory design</td>
<td>2</td>
<td>CYP and researchers</td>
<td>Computer science</td>
<td>2-3 months</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Cassidy et al [51], 2015</td>
<td>Participatory design</td>
<td>6</td>
<td>CYP and researchers</td>
<td>Health</td>
<td>&lt;1 day</td>
<td>7.5</td>
<td>29</td>
</tr>
<tr>
<td>Cheng et al [52], 2018</td>
<td>Participatory design</td>
<td>2</td>
<td>CYP, designers and developers, other experts, and researchers</td>
<td>Serious games</td>
<td>&gt;1 year</td>
<td>20.5</td>
<td>14</td>
</tr>
<tr>
<td>Christie et al [53], 2019</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, clinicians, designers and developers, and researchers</td>
<td>Interaction design</td>
<td>&lt;1 day</td>
<td>Adolescents</td>
<td>30</td>
</tr>
<tr>
<td>de Jans et al [54], 2017</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, teachers, support group, and researchers</td>
<td>Computer science</td>
<td>4-6 months</td>
<td>15</td>
<td>109</td>
</tr>
<tr>
<td>Dürl et al [22], 2017</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, teachers, support group, and researchers</td>
<td>Serious games</td>
<td>1-7 days</td>
<td>11.5</td>
<td>58</td>
</tr>
<tr>
<td>Eriksson et al [55], 2019</td>
<td>Co-design</td>
<td>4</td>
<td>CYP, other experts, designers, and developers</td>
<td>Computer science</td>
<td>&lt;1 day</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Study</td>
<td>What method or methods were labeled or framed?</td>
<td>Participant groups, n</td>
<td>Groups involved</td>
<td>Discipline</td>
<td>How long was the study?</td>
<td>Median age (years)</td>
<td>Young people, n</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Gennari et al [56], 2019</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, researchers, and other experts</td>
<td>Education</td>
<td>1-7 days</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>Gonsalves et al [57], 2019</td>
<td>Person-centered approach</td>
<td>3</td>
<td>CYP and teachers</td>
<td>Game design</td>
<td>4-6 months</td>
<td>10.5</td>
<td>46</td>
</tr>
<tr>
<td>Kang et al [58], 2021</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, teachers, and researchers</td>
<td>Health</td>
<td>2-3 months</td>
<td>9.5</td>
<td>7</td>
</tr>
<tr>
<td>Kangas [59], 2010</td>
<td>Design-based research</td>
<td>3</td>
<td>CYP, carers and family members, and researchers</td>
<td>HCI</td>
<td>&lt;1 day</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>Khaled and Vasalou [60], 2014</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, teachers, and researchers</td>
<td>Computer science</td>
<td>2-3 months</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Kostenius et al [61], 2018</td>
<td>Participatory design</td>
<td>5</td>
<td>CYP, carers and family members, teachers, and researchers</td>
<td>Psychology</td>
<td>N/A</td>
<td>10.5</td>
<td>18</td>
</tr>
<tr>
<td>Lee et al [62], 2019</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, teachers, and researchers</td>
<td>Psychology</td>
<td>2-3 months</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Leitao et al [63], 2019</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, clinicians, and researchers</td>
<td>Education</td>
<td>N/A</td>
<td>9.5</td>
<td>36</td>
</tr>
<tr>
<td>Malinverni et al [64], 2014</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, carers and family members, designer and developer, teachers, and researchers</td>
<td>HCI</td>
<td>7-9 months</td>
<td>7.5</td>
<td>4</td>
</tr>
<tr>
<td>Martens et al [65], 2018</td>
<td>Co-design</td>
<td>2</td>
<td>CYP, teachers, designers and developers, clinicians, and other experts</td>
<td>Computer science</td>
<td>2-3 months</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Marti et al [66], 2016</td>
<td>Co-design</td>
<td>2</td>
<td>CYP, carers and family members, teachers, and support group</td>
<td>Game design</td>
<td>2-4 weeks</td>
<td>9.5</td>
<td>6</td>
</tr>
<tr>
<td>Metatla et al [67], 2020</td>
<td>Co-design</td>
<td>3</td>
<td>CYP, teachers, and researchers</td>
<td>Interaction design</td>
<td>&gt;1 year</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Nouwen et al [68], 2016</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, other experts, designers, and developers</td>
<td>Computer science</td>
<td>10-12 months</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>Patchen et al [69], 2020</td>
<td>Participatory design</td>
<td>3</td>
<td>CYP, designers and developers, and researchers</td>
<td>Health</td>
<td>N/A</td>
<td>16.5</td>
<td>86</td>
</tr>
<tr>
<td>Pavarini et al [70], 2020</td>
<td>Participatory design</td>
<td>4</td>
<td>CYP, researchers, designers, and developers</td>
<td>Education</td>
<td>&gt;1 year</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>Pollio et al [71], 2021</td>
<td>Participatory design</td>
<td>2</td>
<td>CYP and researchers</td>
<td>HCI</td>
<td>N/A</td>
<td>6.5</td>
<td>13</td>
</tr>
<tr>
<td>Porcino et al [72], 2015</td>
<td>Participatory design</td>
<td>4</td>
<td>CYP, carers and family members, and clinicians</td>
<td>Game design</td>
<td>4-6 months</td>
<td>9.5</td>
<td>6</td>
</tr>
<tr>
<td>Powell et al [73], 2019</td>
<td>Realist evaluation</td>
<td>5</td>
<td>CYP, teachers, and researchers</td>
<td>Interaction design</td>
<td>2-3 months</td>
<td>13.5</td>
<td>7</td>
</tr>
<tr>
<td>Rötökönen et al [74], 2021</td>
<td>Co-design</td>
<td>3</td>
<td>CYP and researchers</td>
<td>Education</td>
<td>N/A</td>
<td>9.5</td>
<td>5</td>
</tr>
</tbody>
</table>

https://games.jmir.org/2023/1/e42680 JMIR Serious Games 2023 | vol. 11 | e42680 | p.568 (page number not for citation purposes)
Raynes-Goldie and Allen [28], 2014

Regal et al [75], 2020

Romero et al [76], 2018

Stalberg et al [77], 2016

Sutton et al [78], 2020

Terlouw et al [79], 2021

Triantafyllakos et al [80], 2011

Vallettin-Holbech et al [81], 2020

Vasalou et al [82], 2012

Waddington et al [83], 2015

Werner-Seidler et al [84], 2017

Zhu et al [85], 2019

The youngest age sampled across all studies was 3 years, and the oldest was 25 years as a part of a “young people” sample spanning ages 16 to 25 years [52]. The median age sampled was 11.5 years. Most studies (30/47, 64%) sampled an age range from 0 to 3 years (where 0 would be a precise age in years and 3 would be an age range of, eg, 6 to 9 years; Table 3).

Although our review focused on studies involving children, we were also interested in the participant groups that were involved. The plurality of studies (23/47, 49%) involved 3 different participant groups. The predominant groups were CYP (n=47, 100%) owing to inclusion criteria specifying CYP involvement. Studies also label CYP as patients, learners, or target audiences.
Subject-matter experts (39/47, 83%), such as health professionals (11/47, 23%), teachers (18/47, 38%), and other experts (10/47, 21%), were cumulatively the second-most frequent participating group (Table 4). Other experts encompassed were from areas such as film, photography, art, and music. The researchers conducting or facilitating the study were the next most frequently involved group (36/47, 77%) in the design process. Researchers included those carrying out the study and references to scientists who were also involved in the study. Some studies also used designers and developers (17/47, 36%), which included game design companies, masters’ students in HCI, game designers, graphic designers, and artists, to help with the development of games or interactive technologies. Support groups (5/47, 11%) included learning assistants, pastoral care coordinators, and special education coordinators who helped support CYP participation.

Study length varied greatly, and in 17% (8/47) of studies, the study length could not be determined from the sampled records. The most frequent duration was 2-3 months (10/47, 21%), followed by studies conducted in ≤1 day (8/47, 17%). Studies conducted in ≤1 day were often short workshops or sessions of 1- to 2-hour activities. Studies conducted over 1 to 7 days (5/47, 11%) were the next most frequent. This was followed by studies that took either 4-6 months (4/47, 9%) or 10-12 months (4/47, 9%). Studies conducted between 2 and 4 weeks (3/47, 6%) and 10-12 months (4/47, 6%) were less frequent, and the lowest frequency was 7 to 9 months (2/47, 4%). A consideration to take into account is that it is not clear whether some studies are reporting the length of the whole study, including recruitment, procedure, and analysis, or whether they are reporting the duration of participants involvement in the study.

**Table 3.** Age variance (n=47).

<table>
<thead>
<tr>
<th>Age range variance (years)</th>
<th>Frequency of studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7 (14)</td>
</tr>
<tr>
<td>1</td>
<td>5 (11)</td>
</tr>
<tr>
<td>2</td>
<td>10 (21)</td>
</tr>
<tr>
<td>3</td>
<td>8 (17)</td>
</tr>
<tr>
<td>4</td>
<td>6 (13)</td>
</tr>
<tr>
<td>5</td>
<td>2 (4)</td>
</tr>
<tr>
<td>6</td>
<td>3 (6)</td>
</tr>
<tr>
<td>7</td>
<td>3 (6)</td>
</tr>
<tr>
<td>8</td>
<td>0 (0)</td>
</tr>
<tr>
<td>9</td>
<td>3 (6)</td>
</tr>
</tbody>
</table>

**Table 4.** Stakeholder groups involved (n=47).

<table>
<thead>
<tr>
<th>Stakeholder group</th>
<th>Frequency in studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children and young people</td>
<td>47 (100)</td>
</tr>
<tr>
<td>Researchers</td>
<td>36 (77)</td>
</tr>
<tr>
<td>Teachers</td>
<td>18 (38)</td>
</tr>
<tr>
<td>Designers and developers</td>
<td>17 (36)</td>
</tr>
<tr>
<td>Health professionals</td>
<td>11 (23)</td>
</tr>
<tr>
<td>Other experts</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Carers and family members</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Support groups</td>
<td>5 (11)</td>
</tr>
</tbody>
</table>

**User Involvement Methods Used**

**Self-labeled User Involvement Method**

The overwhelming majority of studies self-labeled their user involvement method as “participatory design” (20/47, 43%) or “co-design” (16/47, 34%; Table 5). However, different studies have used and understood these terms differently with no stable consensus. Some considered participatory design as the overall research area and co-design as the method [56,60,77,82]. Others positioned the 2 as separate methods [62,66], while others used terms interchangeably [75,85]. The most common definition of “participatory design” was to “involve end users in the design process” [45,48,68,72,86], which some interpreted strongly because end users fully and equally participated throughout the whole design and development process [68,70], while others read it weakly as “invit[ing] users to contribute ideas” [71]. There was further less consensus and clarity regarding the meaning and use of “co-design.”
Table 5. Frequency of self-labeled user involvement methods (n=47).

<table>
<thead>
<tr>
<th>User involvement method labeled</th>
<th>Frequency in studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participatory design</td>
<td>20 (43)</td>
</tr>
<tr>
<td>Co-design</td>
<td>16 (34)</td>
</tr>
<tr>
<td>Cocreation or cocreativity</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Design-based research</td>
<td>2 (4)</td>
</tr>
<tr>
<td>User-centered design</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Realist evaluation</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Iterative design</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Participatory action research</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Person-centered approach</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

**Stated Goals of User Involvement**

Several studies have expressed >1 goal of involving users (Table 6). The most frequently stated aim was to design a game, either a prototype or a finished system (22/47, 47% of studies); 38% (18/47) of studies aimed to gather feedback on features and functionality; 26% (12/47) of studies were “meta”-studies aimed at examining involvement, that is, understanding the impact and form of CYP involvement overall; and a further 21% (10/47) of “meta”-studies explored a particular user involvement method or technique. For example, Benton and Johnson [48] explored the use of the participatory design approach to meet the needs of young people with autism spectrum disorder. A further 19% (9/47) of studies aimed to develop skills, meaning that the process served as a learning outcome, such as developing CYP design skills. Overall, 17% (8/47) of studies aimed to understand CYP needs and preferences around gaming, and a further 15% (7/47) of studies aimed to understand CYP’s perception and concerns around the context or technology of the study. Overall, 6% (3/47) of studies focused on improving the user experience of existing products or prototypes, and finally, 4% (2/47) of studies explicitly aimed to create guidelines on how to conduct co-design with CYP (other studies did produce recommendations or guidelines but did not state this as the intended goal behind user involvement [47,62,64,73,82]).

In summary, the included studies stated a wide range of reasons for involving users, with the overwhelming majority focusing on directly informing design and development, from formative research on user needs, preferences, concerns, and contexts (15 studies) to feedback on features and functionality (18 studies), directly making a game (20 studies), or improving user experience (3 studies). Against this stand comparatively fewer “meta”-studies in user involvement and supporting methods themselves (22 studies) and the aim to skill up CYP (9 studies).

Table 6. Frequency of user involvement goals (n=47).

<table>
<thead>
<tr>
<th>Goals of studies</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design a game with participants</td>
<td>22 (47)</td>
</tr>
<tr>
<td>Feedback on features and functionality</td>
<td>18 (38)</td>
</tr>
<tr>
<td>Examine involvement</td>
<td>12 (26)</td>
</tr>
<tr>
<td>Explore methodology</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Develop skills</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Understand children and young people needs and preferences</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Understand perceptions and concerns</td>
<td>7 (15)</td>
</tr>
<tr>
<td>Improved user experience</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Create guidelines</td>
<td>2 (4)</td>
</tr>
</tbody>
</table>

**Involvement Roles of CYP**

The dominant verbatim labels used for CYP involvement roles were “informant” (26/47, 62% of studies) and “co-designer” (20/47, 42%), followed by “playtester/tester” (4/47, 9%), “validation” (4/47, 9%), “co-creator” (4/47, 9%), “end user” (2/47, 4%), and “co-researcher” (1/47, 2%). Overall, 11% (5/47) of studies did not indicate a particular role of involvement (note that a single study could identify multiple roles for CYP; hence, percentage added up to >100%).

Similar to self-labeled methods, these role labels can cover a wide variety of actual degrees and types of involvement. In addition, we coded all studies using the Druin [39] influential taxonomy of 4 possible roles children can play in the design of new technology, in which each successive role can be seen to be more agentic than and encompass the previous ones: (user]tester]informant]design partner!). Each role also differs in the CYP-adult interaction (from indirect observational input to feedback, dialogue, and elaboration on adult ideas), forms of technology materializations engaged with (from ideas to...
prototypes to existing products), and goals of inquiry (from developing theory about use to assessing technology effects to improving design and usability):

1. **CYP as users** describes studies in which adults observe CYP using technology to assess technology effects and build use theory.

2. **CYP as testers** describes studies where CYP act as usability (or play-) testers, usually of prototypes, which can help assess effects, build theory, and improve designs; here, CYP can provide direct feedback on designs.

3. **CYP as informants** describes studies in which CYP can be involved in the full spectrum of human- or user-centered design methods, from formative research, such as interviews, contextual inquiries, and the like, to testing and validating prototypes at the evaluative end; here, CYP can engage in dialogue with adults and elaborate on their ideas and concepts.

We coded each included study according to these roles by examining how CYP had been involved in the procedure of the user involvement method and mapping this to the Druin [39] taxonomy (Table 7). According to this classification, in 51% (24/47) of the studies, CYP were involved as design partners; 70% (33/47) of studies involved CYP as informants; and testers and users were involved in 11% (5/47) of studies each. In 2 studies, CYP were only involved as users, and 1 study involved them only as testers. Finally, it is worth noting that 2 studies explicitly adopted Druin [39] taxonomy and ensured the involvement of CYP in each of the 4 roles [43,75].

### Table 7. Role of children across studies.

<table>
<thead>
<tr>
<th>Role</th>
<th>Total number of occurrences</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>5</td>
<td>[43,58,66,72,75]</td>
</tr>
<tr>
<td>Tester</td>
<td>5</td>
<td>[28,43,53,62,75]</td>
</tr>
<tr>
<td>Informant</td>
<td>33</td>
<td>[41,43,45,46,48-50,52,53,55-57,59,60,63,64,66-70,74-84]</td>
</tr>
<tr>
<td>Design partner</td>
<td>24</td>
<td>[22,43,44,46,47,49,51,54,55,59-63,65,68,69,71,73,75,79-81]</td>
</tr>
</tbody>
</table>

### How User Involvement Is Implemented in Detail

Inductive coding of the studies’ method descriptions yielded 3 high-level categories of how user involvement was implemented in detail: structure and sessions, activities, and media and tools. *Session structure* describes recurring stages and an overarching facilitation organization. *Activities* captures the specific tasks conducted with stakeholders, while *media and tools* describe the range of media and existing games used in activities.

#### Session Structure

In total, 8 studies involved an initial *onboarding or sensitizing stage* to create familiarity with the project topic and team, communicate its goal, and help in understanding the upcoming process [22,50,54,68,74,75,78,81]. This was suggested to build trust between participants, foster user engagement, and make an effective co-design process more likely [22,81]. In 2 studies, onboarding consisted of exploring an existing solution to familiarize oneself with the technology and underlying concepts of the research area [48,81]. In another study, the sensitization steps involved introducing users to the topic through challenges and competitions, which also helped build trust with facilitators [22].

Ten studies [41,42,44,47,48,59,63,64,82,85] described an *ideation or brainstorming stage* to help frame user needs and provide users with a starting place. Notably, brainstorming could be used to generate game ideas [42] or to understand existing user practices and areas of technology-based support [44]. This was often supported by starting exemplars and paper templates [41,47,59,63,64], such as empty scenario storyboards or empty mobile phone screens, or a homework task and prepared video-seeding material for ideas [50]. Two repeatedly mentioned challenges and considerations for this phase are the common “groupthink” converging of participant groups on a first or loudly voiced idea blocking further ideation [47,48] and CYP capabilities to actually conduct ideation [47]—although some noted that CYP tends to bring a beneficial high degree of gaming literacy [82].

In total, 17 studies [45,47,49,51,52,55,56,58,62,63,65,70,71,74,75,77,79] reported the *prototyping stage*. These prototyping stages included sessions in which CYP were involved in the design and development of game ideas [52,58,75], game characters or narrative [49,62,71], and generated alternative ideas to an existing idea [63,77]. Prototyping was seen to afford a sense of ownership in the resulting designs for CYP [62], often challenging researchers’ assumptions on end users, and often leading to in-depth reflection from participants [55,70,71]. Reflection was 2-fold, in which studies designed sessions of reflection, and reflection surfaced unexpectedly. For example, 2 studies designed reflection meetings to iterate and improve future design sessions [45,74], whereas another study found that the design process led to participants reflecting on their game-play experiences compared with other participants [55].

One study deliberately scheduled reflection meetings after prototyping as the basis for future iterations [74]. Two challenges observed in this phase were the limitations of the prototyping tools used and what CYP wanted to portray, in which the limitations of paper prototyping were challenging for CYP to articulate the actions they expected from a digital prototype [51].

An *evaluation stage* was rarely mentioned across studies, where the design process was discussed with participants on its engagement, effectiveness, and efficacy as a user involvement...
method. Outside the reflection meetings, 4 studies scheduled an evaluation stage to gather feedback on the user involvement process [45,47,80,81]. These studies discussed how to make the interaction during the design process more engaging [45], how they found the process as a whole [80], how they could be more involved [47], and how they could cooperate with others [81]. Notably, only 7 studies evaluated the prototypes developed at the end of the study [43,45,47,52,53,80,81].

Activities

Our coding resulted in 45 different activities, of which 19 (42%) were only shown in a single study (omitted in Table 8). Paper prototyping was the most frequently conducted activity, mentioned in 20 studies with labels, such as “paper play activity” [44] or “sketchbook prototyping” [52]. We coded this separately from low-fidelity prototyping (used in a further 7 studies); even though the boundaries between the 2 are not clear-cut, some low-fidelity prototyping would involve paper storyboards and sketches. Regardless, both were reported as affording a positive experience to end users [58,68,80], for example, by giving every participant some hands-on experience [45]. Hands-on experience affected both agencies, in which hands-on experience was a method to assess games with users and understand outcomes [65,66], and learning, in which hands-on experience served as a method of learning technology or understanding the context [28,45,59]. The prepared templates were repeatedly mentioned to facilitate prototyping [63,86]. Paper prototyping was reported to be inclusive [45], low cost [51,65], and using only easily accessible materials [51]. Observed challenges in prototyping included struggling to represent the intended playful interaction with digital technology [51], and that was less suited to older teens because of the hypothetical or “blue-sky” situations when they are at a developmental point of building their own opinions distinct from others [49].

Mapping between activities and design stages was difficult because activities were used across different design stages and in cases that were not transparent when and why activities were used. For example, (focus) group discussions, the second-most prevalent activity (19 studies), were used as icebreakers [45,56], to generate ideas [46,50,63] and reflect on the end product [22,55,57,61]. In another instance, the fourth-most prevalent activity interviews (used in 17 studies) were used equally formatively to discover and define the problem space [41,59,82] and evaluate prototypes or concepts [24,52,53,81]. Presumably because our study sample overall leaned toward “earlier” sensitizing, ideation, and prototyping stages, most activities were used in these stages; only feedback sessions, some instances of game-play evaluation, and 1 timeline activity [68] (asking CYP players to chart their game-play likes and dislikes and experiences of challenge) occurred during an evaluation phase.

The included studies entailed little explicit reflection or evidence of the effectiveness of the conducted activities, with a few exceptions. For example, Nouwen et al [68] outlined which particular activities generated particular user insights and related design impacts. Pavarini et al [70] organized feedback sessions in which CYP could suggest features and processes for better future user involvement. However, even these observations remain unvalidated and can disagree with one another. Thus, although several studies recommended “free play” to provide CYP creative freedom of expression [55,68,86], Nouwen et al [68] found that this had little design impact because the media created during free play were unsuitable for the design brief. Further details of methods and activities can be found in Multimedia Appendix 3.
Table 8. Frequency of activities across studies that were used in more than one study (n=47).

<table>
<thead>
<tr>
<th>Activities</th>
<th>Description</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing games (demonstration or play)</td>
<td>Participants are given existing games, game prototypes, or materials to play or interact with</td>
<td>18 (38)</td>
</tr>
<tr>
<td>Group discussion or focus group discussion</td>
<td>Participants work together and talk about ideas, concepts, or solutions</td>
<td>19 (40)</td>
</tr>
<tr>
<td>Paper prototyping</td>
<td>Participants use mainly paper to design ideas and solutions</td>
<td>20 (43)</td>
</tr>
<tr>
<td>Interviews (with end users)</td>
<td>Researchers interview end users (CYP&lt;sup&gt;a&lt;/sup&gt;) for their thoughts, preferences or feedback on a prototype, context, or problem area</td>
<td>17 (36)</td>
</tr>
<tr>
<td>Playtesting</td>
<td>Participants are given a prototype game and asked to play it (feedback is optional sometimes)</td>
<td>14 (30)</td>
</tr>
<tr>
<td>Storyboarding</td>
<td>Participants are given panels, comic strips, and slides and asked to generate how they would use the game, a scenario, or a solution</td>
<td>11 (23)</td>
</tr>
<tr>
<td>Surveys or Questionnaires</td>
<td>Both qualitative and quantitative surveys. Includes tools such as Likert scales and researchers directly asking CYP to give an answer on a scale</td>
<td>12 (26)</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>Participants work together to generate ideas through group discussions, activities, or in some cases, individually</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Role-play (or simulation)</td>
<td>Participants take on the role of others, for example, CYP acting as doctors, to bridge the challenges different stakeholders undergo</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Feedback Session</td>
<td>Participants are shown a prototype and asked to give feedback</td>
<td>7 (15)</td>
</tr>
<tr>
<td>Introductory media (presentation, movie, etc)</td>
<td>A presentation, movie, talk, or activity is used to help onboard young people on what the goal of the workshop or activity is</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Low-fidelity prototyping</td>
<td>Prototyping with models, figures, blocks, and early development digital games</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Free play</td>
<td>Participants are given a game or demo and no goals are set; they can interact how they choose</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Sticky notes</td>
<td>Tasks involving sticky notes or post its, usually a group-driven task</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Scenario-based tasks</td>
<td>Participants are given a scenario in relation to the applied content with which to design a solution</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Blank template task</td>
<td>Participants are given a template, for example, the wireframe of a phone screen, and use these to design solutions</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Game idea or feature evaluation</td>
<td>Specifically when participants evaluate a component or proposed idea of a game</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Logbooks (or taskbook or diaries)</td>
<td>Participants complete a workbook or taskbook after activities; also includes participants keeping a log of their behavior</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Storytelling</td>
<td>Users create and share stories or concepts</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Timeline (narrative design)</td>
<td>Creating a journey, or series, of steps in which a procedure, process, or story occurs over time or between different users</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Game design workshops</td>
<td>Workshop that encompassed ideation and then evaluation and reflection of those ideas on how they can be improved</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Icebreakers</td>
<td>Informal discussions and activities for participants to get to know each other</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Map task</td>
<td>Using real world maps to choose a setting for their game idea and generate ideas [41,66]</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Peer interviews</td>
<td>Interviews conducted by CYP to other CYP to gather data and help understand the context area</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Question and answer</td>
<td>An activity were users presented predetermined answers (eg, cards) to questions proposed by researchers</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Sensitization session</td>
<td>“Packages” or information circulated before user involvement sessions to understand CYP experience with context and games</td>
<td>2 (4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>CYP: children and young people.
Media and Tools

Following the study by Brandt et al [87], participatory design tools and media can be classified into whether they support the practices of making, telling, and enacting. Bossavat and Parsons [50] observed that 2D mapping (making tool) and videos (telling tool) were useful for understanding concepts, while playing games (enacting) supported idea generation.

The choice of the media or tools used was a point of contention. Some suggested that CYP struggled to express themselves speaking while at the same time preferred speaking to writing [45,47]. In comparison, visual approaches and working with visual aids proved more engaging and effective, especially in younger age groups, potentially owing to their less-developed reading and writing abilities [49,50,62,66]. An exception to the method of media was seen in the studies by Metatla et al [67] and Regal et al [75] with visually impaired children, in which educational robot toys and physical building blocks were used to facilitate creativity through touch and sound. Overall, it seems that preferences for spoken, written, or visual media are related to the age of participants [45,49,62]. Therefore, this suggests that the choice of media and tools used should be determined by understanding participant preferences before the activities take place.

In total, 18 studies used existing games as activities. Existing games varied between existing prototypes developed before the study and existing commercial games [71,76,88]. Existing games were used not only as a tool for narrowing the scope of the research and “managing the expectations” of young participants but also as an icebreaker activity [76,88]. Existing prototypes were sometimes used as a starting point for prototyping or were introduced to users for feedback [45,47,76]. That said, several studies found that concepts and features proposed by CYP were usually informed by the commercial games they were familiar with [50,60,82,85].

Factors Affecting User Involvement

Overview

In this section, we report 5 prominent themes that studies have repeatedly covered as affecting user involvement: comprehension, cohesion, confidence, accessibility, and time constraints. We intentionally did not tie these themes to one particular conception or standard of “good” involvement. Rather, we took the study authors’ own conceptions at face value: if a study articulated something as problematic or positive, we took it as such, also to reflect the variety of goals studies articulated for user involvement.

Comprehension

This theme captured that CYP regularly struggled to understand the design process and context they participated in. CYP repeatedly reported that they did not fully comprehend or remember the benefits or outcomes of the project and how it would impact their daily life [44] nor the aim or purpose of the task assigned to them [72,77,81]. A usability evaluation of one involvement method similarly revealed that tasks were either hard to identify and understand or hard to perform [77]. This problem could be even more common, as Waddington et al [83] observed that CYP did not mention when they struggled with an activity or prototype in their study, so long as they could engage with it; they therefore found it often necessary to gather additional external stakeholder feedback, because CYP did not complain. According to Porcino et al [72], some of these comprehension issues could arise from the lack of a clearly stated objective, insufficient time for a task or technology to become familiar with it, or facilitators not familiar with a given technology or method.

Relatively, how well-informed CYP were about a subject matter impacted the efficacy and efficiency of user involvement: participants who were familiar with a given context proved to be more productive and gave much more concise feedback than unfamiliar ones [53]. Similarly, studies have found it difficult to define problems and identify designs in areas where participants were ill-informed about [54,81]. This should not be considered a one-way street. For example, Durl et al [22] found that involving vulnerable adolescents in a co-design study on alcohol abuse not only resulted in better design results but also made the adolescent participants more informed about alcohol abuse.

A final challenge to comprehension was lacking familiarity with used technologies and methods: this could be the use of basic technology like microphones [74] or the fact that participatory methods differ strongly from what CYP may be used to do in adult-guided activities in school. CYP expected asymmetry in power with adults dictating the direction of the study but experienced more symmetry in power and decision-making because of their suggestions being acknowledged [81]. This challenge could be addressed with additional introductory training and warm-up activities, which consume additional money and time [60].

Cohesion

Most studies involved ≥3 different groups of stakeholders. Therefore, it was unsurprising that many studies reflected on different groups working together effectively as a united team as a major factor of effective user involvement, which we here will call cohesion.

First, several studies reported struggling to achieve agreement on a concept, solution, or decision, likely because of the control each stakeholder group was given [42,74,76]. Control in terms of decision-making and contributing to the design process could be caused by expecting CYP to come to a natural agreement without clear constraints or guidance from adults. Two studies in particular presented open questions to young people and then presents “no right answer,” which could be the cause of friction between young people [74,76]. Common approaches to this issue were discussion (for understanding different points of view) and voting mechanisms (for integrating disagreeing points of view into a decision) [76]. A connected challenge was ensuring that different end-user groups contributed more or less equally to decisions and end results [42,50,64,65]. Several authors observed that cohesion required trust between participants, although no particular approach has been suggested to build trust [42,76].
Another key factor for cohesion was a cooperative and collaborative mindset and atmosphere. Seven studies noted that these allowed CYP to express their views, support each other, and share ideas [22,28,42,55,59,80,81]. In contrast, a competitive dynamic was found to result in less-open idea sharing [76]. The study by Triantafyllakos et al [80] was one of the only studies to reference incentives, in which they used competition and challenges among participants to generate more ideas.

Breakdown in communication could lead to frustration and disappointment [76]. So, how can we afford to function in communication? One study found that a “natural” flow of communication among CYP in which each would have a chance to speak required dedicated facilitation [74], whereas another found that CYP compared with adults were much more direct and unfiltered in criticism, suggesting adult stakeholders should capitalize on this and not be guarded or protective of their ideas [47]. Using existing examples of games [45], as well as explicit tools and conceptual frameworks [55], was found to facilitate discussion.

**Confidence and Empowerment**

Confidence describes CYP’s beliefs in their ability to effectively participate in user involvement, akin to the psychological construct of self-efficacy [89], and influences their participation in several ways [43,68].

CYP would participate more deeply if they felt more confident, which was largely seen as grounded in their past experience; for example, CYP’s past experience with film production or game-making [81], using existing technology CYP are familiar with [44,85], immediate past experience of progress in particular workshop activities [68], or simply longer participation over time [47] would all increase CYP confidence, with positive effects.

However, studies differed in their view of whether CYP are generally confident to voice their opinions [43,83]. Two studies found that CYP had no apparent issue voicing their views directly [47,53], whereas others suggested to use behaviorally “honest signals,” such as eye tracking, because CYP aged 5 to 12 years may say what adults like to hear [43,76] or observed parental interference as an obstacle to CYP sharing unfiltered feedback around sensitive topics like sexual health [62]. Flexibly adjusting group sizes and session lengths to fit CYP needs was found to make them more comfortable overall and share their views more openly [86].

Empowerment was also discussed in studies where young people were given control to make decisions or choices, facilitated by participating in activities, ownership of end products, and the innate challenge of designing a game, all contributed to a sense of empowerment [45,61,81]. While giving CYP power to make decisions over the end product increased a sense of empowerment [77,86], multiple and repetitive user involvement activities were reported to diminish it over time [64]. In addition, there were concerns about the misuse of empowerment as a method of manipulating participants to support the findings desired and empowerment tokenism, where the control over things lacks importance and others make important decisions [61]. Treating CYP as true design partners in game design was found to make it easier for adult participants to connect with CYP’s concerns and, in turn, foster creativity [28].

**Accessibility**

Accessibility describes whether the CYP felt the involvement process was accessible to them and included them in the design. Accessibility was mentioned as a concern, especially in studies that targeted an end-user CYP population with specific needs that forced shaping user involvement around them [64,74,79,83]. Tailoring involvement tools for CYP’s varying abilities [50] and engaging experts associated with CYP’s disabilities (such as carers or teachers) [67] were 2 proposed strategies for improving accessibility. One game-specific accessibility issue mentioned was the development of game mechanics that would be accessible to all players, including those with disabilities, but remains challenging for all end-user groups [74,83].

**Time Constraints**

Time constraints have emerged as a major consideration in structuring user involvement. Co-designing games is time-consuming [65], to the point where de Jans et al [54] suggested that the months required for adults and CYP learning how to collaborate does not fit industry game development timelines. Several studies found that practical time constraints resulted in insufficient discussion of all ideas [74], insufficient preparation, and insufficient time for producing deliverables [76]. Three studies reported breaking the design process into separate phases across different days or sessions to maintain interest, attention, and energy as another important time-related constraint [51,75,76].

**Discussion**

**Principal Findings**

Out of a total of 47 studies, 36 (77%) of our sample were self-labeled participatory design or co-design (Table 5). In line with previous reviews in related fields, such as co-design in health [90] or CYP involvement in child-computer interaction [91], these 2 broad labels gloss over a wide variety of actual involvement roles, aims, and activities. Most studies sampled (35/47, 74%) involved CYP as “mere” informants and about half (24/47, 51%) as true design partners (Table 7). These numbers appear to indicate deeper user involvement for CYP than in recent systematic review of general serious game development processes by Maheu-Cadotte et al [92] (who found only 1 in 21 processes involved participants as co-designers). However, because their inclusion criteria did not limit eligible studies to those with explicit user involvement, this is an unequal comparison.

The stated aims for user involvement ranged from instrumental ones of making a game or identifying features and functionality to meta-methodological interests, upskilling, and many others. We counted 45 different activities, with prototyping, interviewing, playtesting, and playing existing games being the most common. Studies commonly reported “early” boarding, ideation, and prototyping phases; we found little mention of “later” development or production stages [64,78,82]. Furthermore, no records linked or referenced openly shared
user-created design materials outside of screenshots and workshop photographs. Studies used a wide variety of tools and media, with some consensus that these should be tailored to the (developmental) preferences and abilities of the involved CYP. The reviewed studies did not discuss users’ accreditation or rewards for their contributions, outside of 1 example, which suggests uncertainty on how to motivate or reward CYP for their participation.

Our qualitative coding of method reflection produced five factors that are likely to affect successful CYP involvement:

1. Comprehension: the better the CYP know and understand the context and subject matter and design process and used methods, tools, and technologies, the better their engagement and results seem to be.
2. Cohesion: engaging and effective CYP involvement depends on different stakeholder groups working collaboratively as one team, with reaching an agreement as a common challenge and facilitating a comparative noncompetitive mindset as an important success factor.
3. Confidence: the more confident the CYP were in their ability to participate, license to participate, and agency granted to them, the deeper they seemed to engage.
4. Accessibility: both age and disabilities or special needs put extra demand on making user involvement accessible to all participating CYP.
5. Time and space: affording creative time and space can help further the discussion of ideas and maintain interest throughout participation.

Reflection

Perhaps the strongest overarching observation when conducting this systematic review was how differently studies understood and used relevant terms such as “co-design” (2 studies could label their activity “co-design” and “focus group” yet do drastically different things); how differently and unevenly studies reported methods; how differently studies implemented these methods; and in how little detail studies actually documented their involvement methods, despite the fact that our review systematically sampled studies that explicitly reported and reflected on their user involvement. For instance, although we tried to map activities onto “where” in a standard design process, following the example in Vandekerckhove et al [20], we were not able to do so because of the low to absent detail of reporting. This is sadly in line with prior findings on co-design methods: studies rarely report sufficient detail to reproduce them, and there is no standard reporting format [90]. This lack of consensus terminology and reproducible method documentation standards hinders replication. This makes systematic assessment and integration of evidence difficult. In addition, it impedes actual know-how flows beyond tacit and in-person sharing between the members of a research group and project.

Unless we have clear, reliable, and reproducible standards for identifying and reporting types or degrees of user roles and identifying and reporting involvement methods, it is difficult to generate a meaningful systematic body of evidence on which roles or methods work better or worse for which groups of participants, contexts, or aims. This may be one reason why only 4 of the sampled studies reported any qualitative empirical evaluation of what worked and what did not with their involvement methods, even though 46% (22/47) of studies stated meta-methodological aims, such as assessing end-user involvement or involvement methods. Consequently, most reflections and derived recommendations on how to conduct user involvement remain speculative. Thus, we cannot conclude with much certainty anything about, for example, which user roles might work “better.” Although there is a normative preference for the “deepest” possible involvement in much of participatory design, several studies in our review suggested that CYP struggled to act as meaningful co-designers in applied game design [76] and were more able to express themselves as play testers than when sketching their own concepts [63]. This may be partially because of the fact that applied game design requires both domain and game design expertise, which CYP likely does not bring, which may be bridged by more careful onboarding activities or involving multiple stakeholder groups with complementary expertise [60]. Although the involvement of young people has been reported to be essential to realize game prototypes, whether CYP has expertise, or even the availability to consistently collaborate, is a challenge for future research [93]. The important point here remains that we are at present not able to answer such questions or advise on “what works best” because of poor reporting.

Limitations

Our sample is obviously limited by its English-only focus and 10-year date range. The literature before 2010 would potentially paint a distinctively different picture, which would lead to divergent conclusions in this review. Study selection and coding were all performed by the primary author alone, meaning that we could not quantify the likely reliability and reproducibility of these steps, for example, intercoder reliability. The overall poor quality of method reporting in the sampled studies and the lack of existing controlled vocabularies or taxonomies of involvement methods means that our coding involved substantial degrees of judgment and construction in categorizing different kinds of involvement activities or classifying studies by user role. This does not impact verbatim extractions of self-labeling or similar, and we believe our overall assessment of the sheer diversity of reported practices is not touched by this.

Future Research

The first obvious direction for future work is to provide a robust shared basis for reporting and assessing user involvement in applied game development (with or without CYP); we need better consensus methods for defining, labeling, and identifying user involvement roles and methods and better guidance and standards on reporting user involvement in sufficient detail for later analysis or replication. Second, on this basis, we can start to analyze and empirically evaluate what kinds of session structures, methods, roles, media, and tools are more engaging in CYP and more effective for different involvement aims. Third, our review suggests that comprehension, cohesion, confidence, accessibility, and time constraints are likely to impact CYP involvement. Here, methodological research can explore whether and how much these factors matter, and then again, which structures, methods, roles, media and tools, and
specific implementations thereof are better suited to support them.

Conclusions
This scoping review aimed to explore how CYP have been involved in the design of applied games. We found that a small range of labels (co-design and participatory design) hid a wide variety of actual involvement methods, aims, structures, roles, and implementations. Comprehension, confidence, cohesion, accessibility, and time constraints emerged as 5 likely nonexhaustive factors affecting effective and engaging CYP involvement. However, the reviewed literature documented its user involvement practices inconsistently and in little detail, and its recommendations for future practices are largely not grounded in robust empirical evaluations of (alternative) involvement approaches. Future work is needed to advance more robust and reproducible documentation of user involvement to enable knowledge sharing, as well as more systematic research on “what works” in user involvement of CYP in applied game design.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Open Science Framework repository.
[DOCX File, 12 KB - games_v11i1e42680_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File, 14 KB - games_v11i1e42680_app2.docx]

Multimedia Appendix 3
Definition of methods labeled and activities.
[DOCX File, 26 KB - games_v11i1e42680_app3.docx]

References


https://games.jmir.org/2023/1/e42680


Abbreviations

- CYP: children and young people
- HCI: human–computer interaction
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RQ:** research question

**START:** State of the Art through Systematic Review

---

Edited by G Eysenbach, N Zary; submitted 14.09.22; peer-reviewed by F Balli, M Kapsetaki; comments to author 17.11.22; revised version received 06.01.23; accepted 24.01.23; published 16.03.23.

**Please cite as:**
Saiger MJ, Deterding S, Gega L
Children and Young People’s Involvement in Designing Applied Games: Scoping Review
JMIR Serious Games 2023;11:e42680
URL: https://games.jmir.org/2023/1/e42680
doi:10.2196/42680
PMID:36928258

©Michael John Saiger, Sebastian Deterding, Lina Gega. Originally published in JMIR Serious Games (https://games.jmir.org), 16.03.2023. 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.
Review

Effectiveness of Digital Health Interventions Containing Game Components for the Self-management of Type 2 Diabetes: Systematic Review

Linda Ossenbrink¹*, MPH; Tina Haase¹*, Dr rer biol hum; Patrick Timpel¹, Dr rer med; Olaf Schoffer¹, Dr rer nat; Madlen Scheibe¹, Dr rer med; Jochen Schmitt¹, MPH, Prof Dr; Stefanie Deckert¹, MPH, Dr rer med; Lorenz Harst¹, Dr rer med

Center for Evidence-based Healthcare, University Hospital and Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

*these authors contributed equally

Corresponding Author:
Lorenz Harst, Dr rer med
Center for Evidence-based Healthcare
University Hospital and Faculty of Medicine Carl Gustav Carus
Technische Universität Dresden
Fetscherstraße 74
Dresden, 01307
Germany
Phone: 49 37133335325
Email: lorenz.harst@tu-dresden.de

Abstract

Background: Games and game components have become a major trend in the realm of digital health research and practice as they are assumed to foster behavior change and thereby improve patient-reported and clinical outcomes for patients with type 2 diabetes.

Objective: The aim of this systematic review was to summarize and evaluate the current evidence on the effectiveness of digital health interventions containing game components on behavioral, patient-reported, and clinical outcomes for patients with type 2 diabetes.

Methods: An electronic search was conducted in MEDLINE and PsycINFO in April 2020; updated in April 2022; and supplemented by additional searches via Google Scholar, Web of Science (which was used for forward citation tracking), and within the references of the included records. Articles were identified using predefined inclusion and exclusion criteria. In total, 2 reviewers independently conducted title, abstract, and full-text screening and then individually performed a critical appraisal of all the included studies using the Cochrane risk-of-bias tool version 2. A consensus was reached through discussion.

Results: Of 2325 potentially relevant titles (duplicates excluded), 10 (0.43%) randomized controlled trials were included in this review. Quality assessment revealed a high risk of bias for all randomized controlled trials except for 10% (1/10), with performance bias due to the lack of blinding being the major source of bias. There is evidence suggesting that digital health interventions containing game components can substantially improve motivation for physical activity (1/1, 100% of the studies dealing with PA motivation), exercise intensity (3/5, 60%), dietary behavior (4/4, 100%), health literacy (1/3, 33%), mental quality of life (2/2, 100%), glycated hemoglobin level (2/6, 33%), BMI (1/3, 33%), fasting plasma glucose level (1/2, 50%), waist circumference (1/1, 100%), and aerobic capacity (1/1, 100%).

Conclusions: Published studies indicated that digital health interventions containing game components might improve health behavior patterns, quality of life, and clinical outcomes in patients with type 2 diabetes. However, the intervention types and outcomes studied were heterogeneous, and study quality was mostly low, which translates to ambiguous results. Future research should focus on sound methodology and reporting as well as on identifying game components that contribute to significant positive effects.

Trial Registration: PROSPERO CRD42020209706; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=209706

(JMIR Serious Games 2023;11:e44132) doi:10.2196/44132

https://games.jmir.org/2023/1/e44132 JMIR Serious Games 2023 | vol. 11 | e44132 | p.584
(page number not for citation purposes)
**KEYWORDS**
diabetes; gamification; digital health; diabetes self-management; mobile phone

**Introduction**

**Background**

According to the 2021 and most recent Diabetes Atlas provided by the International Diabetes Federation, diabetes affects 537 million people worldwide, which equals a share of almost 7% of the world’s total population [1]. Compared with the data provided in the 2019 Atlas, this is an increase in diabetes prevalence of almost 16% within 2 years [1]. The vast majority of people diagnosed with diabetes (95%) live with type 2 diabetes [2,3].

Apart from genetic predispositions and a higher probability of having a diabetes diagnosis at an older age, several risk factors of type 2 diabetes are lifestyle-related, such as physical inactivity; malnutrition; and, correspondingly, overweight and obesity [4]. Therefore, several clinical practice guidelines on the treatment of type 2 diabetes focus not only on pharmacological interventions but also on a comprehensive self-management regimen that includes theory-based behavior change [4-7]. Measures for the latter are provided by the so-called Diabetes Self-Management and Education (DSME) regimen defined by the American Diabetes Association (ADA) include education on symptoms, etiology of and coping with diabetes, the adoption of a healthy (ie, high fiber- and fruit- and vegetable-based) diet, and the uptake of regular physical activity (PA) as a means to achieve weight loss [4,6]. Apart from the education component, continuous monitoring of blood glucose values, food intake, and frequency and intensity of PA is a requirement of DSME [6].

The potential of digital health applications, such as telemedicine, for supporting patients regarding DSME is well documented for patients with a more recent diabetes diagnosis, especially for applications that enable continuous glucose self-monitoring or health care provider feedback on the values recorded [8].

Recently, interest has spiked in digital health applications containing game components as they are expected to offer aid in behavior change [9], which, for the latter mandated by the so-called Diabetes Self-Management and Education (DSME) regimen defined by the American Diabetes Association (ADA) include education on symptoms, etiology of and coping with diabetes, the adoption of a healthy (ie, high fiber- and fruit- and vegetable-based) diet, and the uptake of regular physical activity (PA) as a means to achieve weight loss [4,6]. Apart from the education component, continuous monitoring of blood glucose values, food intake, and frequency and intensity of PA is a requirement of DSME [6].

The protocol for this systematic review was published beforehand in PROSPERO (CRD42020209706) and followed during the conduct of the review. Reporting of the results adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [23].

**Objectives**

Therefore, the questions to be answered were as follows: (1) Do digital game-based interventions have an effect on the health behavior of patients with type 2 diabetes? and (2) Do digital game-based interventions have an effect on clinical outcomes and PROs in patients with type 2 diabetes?

**Methods**

According to theory, either individuals can assess the situation they are in and then develop the most adequate strategy to reach a behavioral aim (assessment) or they can just initiate the behavior for which they strive (locomotion). Locomotion is associated with higher intrinsic motivation, whereas assessment is associated with anticipating failure and, therefore, procrastination [16].

In 2016, a meta-analysis including both types of game-based interventions (educational games and exergames) for patients with diabetes (type 1 and 2) showed no effect on blood glucose values (glycated hemoglobin [HbA1c]) but showed an effect on quality of life, balance, and muscle strength [17]. Cheng et al [18] found that videogames were useful tools for diabetes education independent of diabetes type. However, the analyses are outdated (both searches were conducted in 2014) given the substantial increase in the availability of gamified health interventions, as evidenced by a recent overview of gamified interventions used in the context of diabetes [19], half of which were launched after 2016. The median survival time of reviews is 5.5 years before they are outdated [20]. In addition, patient populations and the measures taken to deal with type 1 and type 2 diabetes differ greatly [7], and so do the effective components of digital health applications [8]. Therefore, targeted game-based interventions are necessary as well. Furthermore, given the complex requirements of DSME, focusing the analysis solely on clinical outcomes and neglecting behavioral outcomes as well as patient-reported outcomes (PROs), as did Martos-Cabrera et al [21] and Kaihara et al [22], is limited in perspective.

**Study Inclusion and Exclusion**

The inclusion and exclusion criteria were defined according to the population, intervention, control, outcome, and study design scheme (Textboxes 1 and 2). For this purpose, digital health interventions containing game components were defined as the intervention group, whereas usual care or the use of digital health interventions without a game component were defined as the control group.
Textbox 1. Study inclusion criteria.

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome</th>
</tr>
</thead>
</table>
| Participants with type 2 diabetes (no age restriction) | Use of digital health applications containing game components identified in previous reviews [12,17] (such as virtual reality, serious gaming, or exergaming) | Use of digital health applications without game components or standard or usual care | Primary outcomes:  
  - Behavioral outcomes such as physical activity or dietary behavior  
 Secondary outcomes:  
  - Patient-reported outcomes such as self-efficacy, patient empowerment, and quality of life  
  - Clinical parameters such as glycated hemoglobin (blood sugar value), BMI, and systolic blood pressure or diastolic blood pressure |

Study design  
- Randomized controlled trials  
- Nonrandomized studies (only when n>10)

Textbox 2. Study exclusion criteria.

<table>
<thead>
<tr>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants without diabetes or with type 1 or gestational diabetes</td>
<td>No treatment or intervention</td>
<td>No treatment or intervention</td>
<td>Neither behavioral outcomes nor patient-reported outcomes or clinical parameters studied</td>
</tr>
</tbody>
</table>

Study design  
- Cross-sectional studies, qualitative studies, reviews, and meta-analyses

Only studies published in English or German were included.

Database Search

An electronic database search was conducted in MEDLINE (via PubMed) and PsycINFO to cover both medical and psychological research. The initial search was conducted in April 2020 and updated in April 2022 with no restrictions on start time. The search string (Multimedia Appendix 1) included terms covering type 2 diabetes (population), including Medical Subject Heading terms and synonyms for game and gaming (intervention) as well as game components, and was piloted in previous research [24]. Population and intervention terms were linked with the operator AND. No restrictions were imposed on the outcome category of the search string to avoid accidentally excluding relevant effects of gamified interventions.

Additional searches were conducted within the references of the included studies (backward citation tracking) as well as on Google Scholar and Web of Science, where publications citing the included studies were checked (forward citation tracking) in July 2022.

Screening and Data Extraction

The screening of relevant records was a 2-step process. First, 2 independent reviewers screened the titles and abstracts of all records found by the database and hand searches. If deemed relevant by at least one reviewer, the full text was assessed for
eligibility by both reviewers as well. The reference manager EndNote (Clarivate Analytics) was used for both screening and duplicate removal.

Data were extracted according to the population, intervention, control, outcome, and study design scheme, aiming at allowing for a comparison of the effects of different gamified intervention types on the aforementioned outcome domains. In addition, the following information was extracted from each study: (1) bibliographic information, (2) population characteristics, (3) the allocation of the study participants to the intervention and control group or control groups, (4) treatment or interventions applied to the control group (as a means to account for plausible confounding factors), (5) inclusion and exclusion criteria of each applicable study (as a means to account for plausible confounding factors), and (6) outcome measures (to inform quality assessment).

The data extraction sheet was piloted by 2 researchers on 2 of the included studies, which were randomly selected, and subsequently slightly adjusted by including the category “outcome measures.” Data extraction was performed using a Microsoft Excel (Microsoft Corp) spreadsheet.

For a visual representation, HbA$_{1c}$ values at baseline and after intervention completion were extracted in percentage or mmol/mol depending on the data available in the included studies, along with the SD. Δ HbA$_{1c}$ was computed, and statistical significance was extracted from the included studies. The threshold for statistically significant effects was set at $P < .05$. No assumptions were made if information was missing; this was labeled as “not reported” instead. All study results were tabulated. Apart from the tabulation and visualization of HbA$_{1c}$ results, the presentation of the results is narrative.

Quality Assessment

The randomized controlled trials (RCTs) found were assessed for study quality by applying the Cochrane risk-of-bias tool version 2 (RoB 2) [25], whereas cohort studies, if included, were assessed using the corresponding Critical Appraisal Skills Programme (CASP) cohort study checklist [26]. Case-control studies were assessed using the corresponding checklist also provided by the CASP [27]. The RoB 2 covers bias within an outcome resulting from the randomization process (selection bias); blinding of participants, assessors, and analysts; deviations from intervention delivery (performance bias); changes in participants’ adherence to the intervention (attrition bias); modalities of outcome measurement; or selective reporting (reporting bias). RoB 2 assessment was performed for each relevant study outcome according to our inclusion criteria. According to the RoB 2 manual, the overall risk of bias within a study was deemed high if 1 study outcome had a high risk of bias. Both records by Höchsmann et al [28,29] were treated as 1 study with several outcomes. CASP checklists for nonrandomized studies cover the same categories except for randomization and blinding while also putting an emphasis on practical implications of the study results. In contrast to the RoB 2, CASP checklists were applied to the entire study instead of to selected outcomes. The RoB 2 deems studies to be at a low risk of bias when a low risk of bias is detected for all relevant domains. Some concerns regarding the risk of bias within an outcome can be assumed when some concerns are raised for at least one domain. A high risk of bias can be assumed when multiple domains raise some concerns or a high risk of bias is detected for at least one domain [25]. With the CASP, overall risk of bias is assessed by answering the following question: “Do you believe the results?”

Quality assessment was performed for the effect of both assignment to the intervention (ie, the intention-to-treat effect) and adherence to the intervention (ie, the per-protocol effect).

All steps of the review—screening of titles and abstracts and full texts, data extraction, and quality assessment—were performed by at least 2 researchers independently (LO, LH, or PT in the first search period and LH and TH in the second search period) to minimize bias. Differences in inclusion, extraction, and quality assessment were resolved through discussion with a third person not involved in the screening process (SD).

Diversions From the Protocol

Contrary to the protocol registered with PROSPERO, research question 2 was adapted so that it clearly addressed PROs. Furthermore, we refrained from performing a search in key journals as these were all listed in MEDLINE.

Results

Results of the Search Process

Both database searches (2020 and 2022) taken together yielded 2325 results. A total of 7 additional potentially relevant publications were identified via additional searches, but excluded after full text assessment. In total, 2087 publications remained after duplicates were removed, 2034 (97.5%) of which were removed after title and abstract screening. The full texts of the remaining 53 publications were assessed for eligibility. This process led to the exclusion of 25% (13/53) of the studies as they addressed a population outside this review’s scope, such as patients with type 1 or prediabetes [30-42]. Another 40% (21/53) of the studies were excluded as the interventions studied did not have a game component [43-63]. A total of 15% (8/53) of the studies were excluded as they used 1-armed designs or did not study any of the prespecified outcomes but solely patient experiences, such as satisfaction with the application [64-71]. A complete list of the excluded studies with reasons for exclusion can be found in Multimedia Appendix 2 [30-71]. Finally, 10 studies were included in the qualitative data analysis [28,29,72-80]. Höchsmann et al [28,29] reported results from the same study in 2 records. The process of study selection is depicted in Figure 1.
Descriptive Statistics of the Populations Studied

All the included studies (10/10, 100%) were RCTs. However, Kempf and Martin [78] and Brinkmann et al [72] applied a crossover design in which participants in the control group received the intervention later during the study period. The study samples ranged from 8 participants [72] to 465 participants [76]. The lowest mean age was 44 (SD 7.9) years [79] and the highest mean age was 68 (SD 5.8) years [73]. Brinkmann et al [72] only reported the age span of the study participants, which was 67 to 75 years. Overall, 533 female participants and 1061 male participants were included. The intervention duration ranged from 30 minutes (exergame) [72] to 36 weeks (self-management app with quiz component) [74]. In total, 30% (3/10) of the RCTs performed follow-up analyses after the intervention period, with follow-up times ranging from 7 days [72] to 48 weeks [77]. A complete overview of the population characteristics of the included RCTs can be found in Table 1.
<table>
<thead>
<tr>
<th>Study, year, title, journal, and country</th>
<th>Study design</th>
<th>Study duration</th>
<th>Follow-up time</th>
<th>Population characteristics</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
</table>
| Brinkmann et al [72], 2017, “Effects of Cycling and Exergaming on Neurorotrophic Factors in Elderly Type 2 Diabetic Men—A Preliminary Investigation”/Experimental and Clinical Endocrinology & Diabetes | Individually randomized crossover trial | 30 min once | None | 8 | Ranging from 67 to 75 | 8 male participants | Not reported | Nonsmoking
| | | | | | | | Absence of diabetic retinopathy, neuropathy, nephropathy, or cardiovascular complications |
| Dugas et al [73]. 2018, “Individual Differences in Regulatory Mode Moderate the Effectiveness of a Pilot mHealth trial for Diabetes Management among older Veterans”/PLOS ONE | Individually randomized controlled trial | 13 weeks | None | 27 | Mean 67.8 (SD 6.1) | Not reported | Veteran patients with type 2 diabetes | Blindness
| | | | | | | | Deafness |
| | | | | | | | Serious mental illness |
| | | | | | | | Homelessness |
| Glasgow et al [74], 2010, “Outcomes of Minimal and Moderate Support Versions of an Internet-Based Diabetes Self-Management Support Program”/Journal of General Internal Medicine, United States | Individually randomized controlled trial | 16 weeks | None | 463 | Mean 58.4 (SD 9.2) | 233 male and 230 female participants | Aged 25 to 75 years | BMI >25 kg/m² |
| | | | | | | | Type 2 diabetes diagnosis |
| | | | | | | | At least one or more risk factor for CVD |
| | | | | | | | Telephone access |
| | | | | | | | Internet access at least twice a week | None |

Exclusion criteria:
- Nonsmoking
- Absence of diabetic retinopathy, neuropathy, nephropathy, or cardiovascular complications

Inclusion criteria:
- Veteran patients with type 2 diabetes
- Affiliated with a Veterans Affairs medical center
- Aged >60 years
- Poorly controlled diabetes (HbA1c >7.9%)

Exclusion criteria:
- Blindness
- Deafness
- Serious mental illness
- Homelessness
<table>
<thead>
<tr>
<th>Study, year, title, journal, and country</th>
<th>Study design</th>
<th>Study duration</th>
<th>Follow-up time</th>
<th>Population characteristics</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grewal et al [75], 2015, “Sensor-Based Interactive Balance Training with Visual Joint Movement Feedback for Improving Postural Stability in Diabetics with Peripheral Neuropathy: A Randomized Controlled Trial”/Gerontology, United States</td>
<td>Individually randomized controlled trial</td>
<td>45 min twice a week for 4 weeks</td>
<td>None</td>
<td>39</td>
<td>Mean 63.7 (SD 8.2)</td>
<td>20 male and 19 female participants</td>
</tr>
<tr>
<td>Höchsmann et al [28,29], 2019, “Effectiveness of a Behavior Change Technique–Based Smartphone Game to Improve Intrinsic Motivation and Physical Activity Adherence in Patients With Type 2 Diabetes: Randomized Controlled Trial”/JMIR Serious Games and “Novel Smartphone Game Improves Physical Activity Behavior in Type 2 Diabetes”/American Journal of Preventive Medicine, Switzerland</td>
<td>Individually randomized controlled trial</td>
<td>24 weeks</td>
<td>None</td>
<td>36</td>
<td>Mean 57 (SD 5.5)</td>
<td>19 male and 17 female participants</td>
</tr>
<tr>
<td>Kempf and Martin [78], 2013, “Autonomous Exercise Game Use Improves Metabolic Control and Quality of Life in Type 2 Diabetes Patients—a Randomized Controlled Trial”/BMC Endocrine Disorders, Germany</td>
<td>Crossover individually randomized controlled trial</td>
<td>30 min per day for 12 weeks</td>
<td>None</td>
<td>220</td>
<td>Mean 62 (SD 11; [intervention group]) and 60 (SD 9; [control group])</td>
<td>119 female and 101 male participants</td>
</tr>
<tr>
<td>Kerfoot et al [76], 2017, “A Team-Based Online Game Improves Blood Glucose Control in Veterans With Type 2 Diabetes: A Randomized Controlled Trial”/Diabetes Care, United States</td>
<td>Individually randomized controlled trial</td>
<td>24 weeks</td>
<td>48 weeks</td>
<td>465</td>
<td>Mean 59.5 (SD 9.9)</td>
<td>28 female and 428 male participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, year, title, journal, and country</td>
<td>Study design</td>
<td>Study duration</td>
<td>Follow-up time</td>
<td>Population characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koohmareh et al [79], 2020, “Effect of Implementing a Mobile Game on Improving Dietary Information in Diabetic Patients”/Medical Journal of The Islamic Republic of Iran, Iran</td>
<td>Individually randomized controlled trial</td>
<td>15 min per day for 6 weeks</td>
<td>None</td>
<td>60 Mean 44.1 (SD 7.9; control group) and 43.9 (SD 9.0; intervention group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maharaj et al [80], 2021, “Comparing Two Commercially Available Diabetes Apps to Explore Challenges in User Engagement: Randomized Controlled Feasibility Study”/JMIR Formative Research, Australia and New Zealand</td>
<td>Individually randomized controlled trial</td>
<td>2 weeks</td>
<td>None</td>
<td>89 Mean 53.2 (SD 11.1; control group) and 52.6 (SD 13.0; intervention group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnin et al [77], 2021, “Impact of a Remote Monitoring Programme Including Lifestyle Education Software in Type 2 Diabetes: Results of the Educ@dom Randomised Multicentre Study”/Diabetes Therapy, France</td>
<td>Individually randomized controlled multicenter study</td>
<td>48 weeks</td>
<td>None</td>
<td>263 Mean 59.6 (SD 9.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2 diabetes diagnosis confirmed by a specialist</td>
<td>Not reported</td>
</tr>
<tr>
<td>Aged &gt;18 years</td>
<td></td>
</tr>
<tr>
<td>Minimal literacy</td>
<td></td>
</tr>
<tr>
<td>Android-run smartphone</td>
<td></td>
</tr>
<tr>
<td>Smartphone skills</td>
<td></td>
</tr>
<tr>
<td>Willingness to participate</td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes diagnosis confirmed by a specialist</td>
<td>Not reported</td>
</tr>
<tr>
<td>Aged &gt;18 years</td>
<td></td>
</tr>
<tr>
<td>Fluent in spoken and written English</td>
<td></td>
</tr>
<tr>
<td>iOS- or Android-run smartphone</td>
<td></td>
</tr>
<tr>
<td>Written consent</td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes diagnosis confirmed by a specialist</td>
<td>Active severe comorbidities</td>
</tr>
<tr>
<td>Aged &gt;18 years</td>
<td>Reduced mobility</td>
</tr>
<tr>
<td>Insulin- and non–insulin-dependent</td>
<td>Eating disorders</td>
</tr>
<tr>
<td>HbA1c of 6.5% to ≤10%</td>
<td>Bariatric surgery</td>
</tr>
<tr>
<td>Active internet connection</td>
<td></td>
</tr>
</tbody>
</table>

Results of the Quality Assessment

Applying the RoB 2, a high risk of bias was detected in 90% (9/10) of the included RCTs [28,73-80]. A total of 10% (1/10) of the studies had a low risk of bias [72]. For 40% (4/10) of the RCTs, multiple reasons for a high risk of bias were detected [73,74,79,80], whereas for 30% (3/10) of the studies, only 1 reason was found [75,77,78]. The allocation sequence was random in all but 10% (1/10) of the cases [79]. Major sources of bias were the blinding of study participants and personnel to the allocation to either the intervention or the control group [73,74,76,77,79,80] and, to a smaller degree, inadequate (ie, nonreliable or nonvalidated) measures for outcome assessment [28,73-75,79,80], both of which translate to detection bias. Most of the included RCTs (7/10, 70%) conducted per-protocol analyses [28,29,72,73,75,77,78,80], whereas another study did
not specify the type of analysis [79]. As such, the effect estimate was potentially biased by dropouts in all but 20% (2/10) [74,76] of the included RCTs (attrition bias). Glasgow et al [74] and Kerfoot et al [76] performed an intention-to-treat analysis, whereas the remaining authors all performed per-protocol analyses except for Koohmareh et al [79], for whom the type of analysis could not be discerned. Bias because of selective reporting was detected by comparing the outcomes described in the Methods sections and those reported in the Results sections [75,76]. The results of the quality assessment can be found in Figure 2.

**Figure 2.** Results of the quality assessment applying the Cochrane risk-of-bias tool version 2 [28,29,72-80].

<table>
<thead>
<tr>
<th></th>
<th>Generation of randomization sequence</th>
<th>Covered assignment to study groups</th>
<th>Blinding of study participants and personnel</th>
<th>Incomplete outcome data</th>
<th>Outcome assessment</th>
<th>Selective reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkmann et al (2017) [72]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Dogas et al (2014) [73]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glasgow et al (2010) [74]</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Gewald et al (2015) [75]</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Kempf and Martin (2013) [78]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Kerfoot et al (2017) [76]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Maharaj et al (2021) [80]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>Turnin et al (2021) [77]</td>
<td>+</td>
<td>-</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Game Types and Components Analyzed in the Included RCTs**

Of the 10 included RCTs, 4 (40%) focused on an exergame [28,29,72,75,78], one (25%) of which reported first behavioral and then clinical results of the same exergame intervention studied with the same population [28,29]. A total of 20% (2/10) of the studies reported on the effects of game components within digital self-management applications for diabetes [73,80]. In 40% (4/10) of the RCTs, the authors analyzed quiz games with DSME content [74,76,77,79]. The digital intervention studied by Brinkmann et al [72] combined an exergame with a game for cognitive problem-solving.

Within the exergames, virtual reality components such as virtual race tracks were used [72,75,78]. Scoring systems awarding trophies to winners or when individualized scores were achieved were used in 60% (6/10) of the interventions [28,29,73,74,76,79,80]. Storytelling features were part of 20% (2/10) of the interventions studied [28,79], and 10% (1/10) of the interventions applied a team-based approach to the game [78].

Among the 4 RCTs in which the authors reported on the matter [72,75,78,79], intervention intensity varied between 30 minutes once (equal to intervention duration) [72], 15 minutes per day for 6 weeks [79], 30 minutes per day for 12 weeks [78] and 45 minutes twice a week for 4 weeks [75].

A complete overview of the intervention types and game components studied in the included RCTs can be found in Table 2.
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Intervention group</th>
<th>Control group</th>
<th>Intervention Duration</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brinkmann et al [72], 2017</td>
<td>Exergaming: interactive video games using Wii Fit Plus combining physical activity and cognitive challenges (n=8) aiming to improve neurotrophic factors</td>
<td>Cycling on stationary bicycle (n=8)</td>
<td>30 minutes</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Dugas et al [73], 2018</td>
<td>Using self-management app with game elements (point reward system for achieving target values in clinical and behavioral outcomes) aiming to improve diabetes outcomes in 4 randomized conditions:</td>
<td>Usual care (n=5)</td>
<td>13 weeks</td>
<td>Not reported</td>
</tr>
<tr>
<td>Glasgow et al [74], 2010</td>
<td>Self-management website allowing for goal setting in medication adherence, nutrition, and exercise, including quiz component (n=169)</td>
<td>Enhanced usual care (automated feedback on health risks based on patient-provided disease-relevant parameters; recommendation of preventive behaviors; n=132)</td>
<td>16 weeks</td>
<td>Not reported</td>
</tr>
<tr>
<td>Grewal et al [75], 2015</td>
<td>Virtual obstacle-crossing task with audio and video feedback on a virtual reality interface</td>
<td>Usual care (n=18)</td>
<td>4 weeks</td>
<td>45 minutes twice a week</td>
</tr>
<tr>
<td>Höchsmann et al [28,29], 2019</td>
<td>Smartphone game “Mission: Schweinehund” based on self-determination theory and including behavioral or motivational elements as well as a storyline where users restore a garden with resources gained through individualized in-game workouts tracked via phone sensors (n=18)</td>
<td>1-time lifestyle counseling and structured exercise plan (n=18)</td>
<td>24 weeks</td>
<td>Not reported</td>
</tr>
<tr>
<td>Kempf and Martin [78], 2013</td>
<td>Wii Fit Plus including balance board (n=120)</td>
<td>Routine care (n=100)</td>
<td>12 weeks</td>
<td>&gt;30 minutes per day for 12 weeks</td>
</tr>
<tr>
<td>Kerfoot et al [76], 2017</td>
<td>Team-based mobile DSME quiz with scoring system plus booklet on civic issues (n=227)</td>
<td>Team-based mobile quiz on civic issues plus DSME booklet (n=229)</td>
<td>24 weeks</td>
<td>Not reported</td>
</tr>
<tr>
<td>Koohmareh et al [79], 2020</td>
<td>Amoo mobile game—glycemic index and calorie training game (n=30)</td>
<td>Educational content similar to the game through a brochure (n=30)</td>
<td>6 weeks</td>
<td>15 minutes per day for 6 weeks</td>
</tr>
<tr>
<td>Maharaj et al [80], 2021</td>
<td>mySugr—self-monitoring and self-management app with nudging and game component (point reward system for certain behaviors; n=44)</td>
<td>Glucose Buddy—self-monitoring and self-management app (n=45)</td>
<td>2 weeks</td>
<td>Not reported</td>
</tr>
<tr>
<td>Turnin et al [77], 2021</td>
<td>DSME software including quiz components (n=135)</td>
<td>Usual care (n=128)</td>
<td>48 weeks</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

*DSME: Diabetes Self-Management and Education.*
Effects of Gamified Health Interventions on Health Behavior

Exercise intensity and adherence were studied in 50% (5/10) of the included studies [28,29,73-75,78]. A total of 40% (4/10) of the studies reported on nutritional behavior [73,74,79,80], and 20% (2/10) of the studies reported on medication adherence [73,74].

Positive effects on behavioral PA outcomes were reported in several of the included RCTs (4/10, 40%). A total of 50% (2/4) of the exergames studied led to an increased intensity of PA (measured as step count, caloric expenditure, and self-report) in the study participants [28,29,78], and so did the quiz component nested within the DSME website studied by Glasgow et al [74]. However, the latter was only effective when the website was supplemented with follow-up calls by study staff as a supportive measure in handling the website and group meetings with other study participants [74]. Only Höchsmann et al [28] reported on PA adherence and found no substantial effect of their exergame on PA adherence, where PA exercises were nested within a coherent narrative. However, the intervention statistically significantly increased intrinsic motivation for PA (P<.001), whereas intrinsic motivation decreased in the control group (1-time lifestyle counseling and structured exercise plan) during the study period. Between-group effects were significant. The authors also found a significant positive relationship between time spent doing the in-game exercises and PA motivation (P=.01) [28].

Dugas et al [73] found no overall intervention effect of any of the intervention arms studied on adherence as a whole (quantified using an in-app rating system; Table 3), but they found a statistically significant interaction effect (in regression analysis) of time spent using the self-management app with game elements and assessment (P=.01). They also found a statistically significant interaction effect of time and locomotion on weekly exercise adherence (P<.05) [73].

In 40% (4/10) of the RCTs, the authors found significant positive effects on nutrition behavior. Glasgow et al [74] reported a significant decrease in fat intake in the intervention group compared with that in the control group (P value for intergroup differences=.006). Koohmaret al [79] found that participants using their educational game on glycemic index and calorie intake paid significantly more attention to the glycemic levels and calorie count of their food (P value for intergroup differences=.001). Maharaj et al [80] showed that participants using the mySugr app with the reward system could significantly reduce their high-fat food consumption, whereas no significant changes were found in the control group using a self-management app without a game component (P value for intergroup differences=.052). Interaction effects of time and assessment as well as of time and locomotion were found by Dugas et al [73] for nutrition adherence (quantified using an in-app rating system) as well. The authors also found a significant positive effect of locomotion on nutritional adherence (P<.05) but none of assessment (P value not reported) [73].

Glasgow et al [74] found no significant effect of the intervention on medication adherence or on eating habits other than fat intake (P value for intergroup differences=.006). Dugas et al [73] found an interaction effect of time spent using the intervention and assessment on medication adherence (P=.04). A total of 40% (2/5) of the exergames studied had no effect on PA intensity [75,78]. Overall, positive effects of digital health interventions with game components could be found on motivation for and intensity of PA as well as on eating habits.
Table 3. Results of the included randomized controlled trials according to all relevant outcomes and overall study quality.

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Outcome (outcome measure)</th>
<th>Results</th>
<th>Study quality according to RoB 2&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Statistically significant</td>
<td>Nonsignificant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clinical parameters:</td>
<td>Clinical parameters:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significantly better lactate values in intervention group (P=.04)</td>
<td>• Insignificant lower heart rate in intervention group (P&gt;.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Heart rate</td>
<td>• Insignificant increase in BDNF in intervention group (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lactate values</td>
<td>• Insignificant increase in VEGF in intervention group (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinical parameters:</td>
<td>• Insignificant increase in IGF-1 in intervention group (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Behavior outcomes:</td>
<td>• Behavioral outcomes:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant positive causal relationship between locomotion and adherence on HbA&lt;sub&gt;1c&lt;/sub&gt; (P&lt;.01)</td>
<td>• No treatment effect of any intervention arm on total adherence (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Adherence to healthy behaviors (glucose, medication, and nutrition tracking entered manually into app; PA&lt;sup&gt;f&lt;/sup&gt; tracked via Fitbit or manually; and recorded app use, all quantified using in-app point rating system)</td>
<td>• Clinical parameters:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant positive causal effect of interaction between time and locomotion for app use, including provider communication features, on adherence (P=.04)</td>
<td>• No significant between-group differences in HbA&lt;sub&gt;1c&lt;/sub&gt; during intervention time (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Behavioral outcomes:</td>
<td>• Clinical parameters:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant positive causal effect of app use, including provider communication features, on HbA&lt;sub&gt;1c&lt;/sub&gt; (P&lt;.01)</td>
<td>• Significantly lower heart rate in intervention group (P&gt;.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinical parameters:</td>
<td>• Insignificant increase in BDNF in intervention group (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant positive causal effect of interaction between time and adherence on HbA&lt;sub&gt;1c&lt;/sub&gt; (P&lt;.01)</td>
<td>• Insignificant increase in VEGF in intervention group (P value not reported)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinical parameters:</td>
<td>• Insignificant increase in IGF-1 in intervention group (P value not reported)</td>
</tr>
</tbody>
</table>

Brinkmann et al [72], 2017

Dugas et al [73], 2018

<sup>a</sup>BDNF, VEGF, and IGF-1 (all measured using an enzyme-linked immunosorbent assay)

<sup>b</sup>Heart rate

<sup>c</sup>Lactate values

<sup>d</sup>HbA<sub>1c</sub>

<sup>e</sup> Regulatory mode (locomotion and assessment scales developed by Kruglanski et al [81])

<sup>f</sup> PA tracked via Fitbit or manually; and recorded app use, all quantified using in-app point rating system
Study quality according to RoB 2

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Outcome (outcome measure)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistically significant</td>
<td>Nonsignificant</td>
</tr>
<tr>
<td>Glasgow et al [74], 2010</td>
<td>Health literacy (brief questionnaire by Chew et al)</td>
<td>• Behavioral outcomes:</td>
</tr>
<tr>
<td></td>
<td>Dietary behavior (“Starting the Conversation Scale” by Amnernan et al)</td>
<td>• Significant improvement in eating habits in intervention group but not in control group ($P$ value for intergroup differences ≤ .001)</td>
</tr>
<tr>
<td></td>
<td>Adherence to medication for diabetes, blood pressure, and cholesterol (medication-taking items of the Hill-Bone Compliance Scale)</td>
<td>• Significant decrease in fat intake in intervention group but not in control group ($P$ value for intergroup differences = .006)</td>
</tr>
<tr>
<td></td>
<td>Fat intake (National Cancer Institute Percent Energy from Fat Screener)</td>
<td>• Significant increase in PA in intervention group but not in control group ($P$ value for intergroup differences = .04)</td>
</tr>
<tr>
<td></td>
<td>Total weekly caloric expenditure (Community Healthy Activities Model Program for Seniors Questionnaire)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• HbA1c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• BMI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Lipid ratio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mean arterial pressure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PROs:</td>
<td>• No significant improvement in medication adherence in intervention group vs control group ($P$ value for intergroup differences = .29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No significant differences in eating habits ($P$ value for intergroup differences = .08), fat intake ($P$ value for intergroup differences $P = .46$), PA ($P$ value for intergroup differences = .63), and medication adherence ($P$ value for intergroup differences = .86) between intervention groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No interaction effect of health literacy on any of the outcomes measured ($P$ not reported)</td>
</tr>
<tr>
<td></td>
<td>Clinical parameters:</td>
<td>• Significant reduction in average center of mass sway area (in degrees) in intervention group but not in the control group ($P$ value for intergroup differences = .04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction in mediolateral center of mass sway area in intervention group but not in the control group ($P$ value for intergroup differences = .008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction of hip and ankle sway in the intervention group but not in the control group ($P$ value for intergroup differences = .008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction of ankle sway degree in intervention group but not in control group (with and without blindfold; $P$ value for intergroup differences = .02)</td>
</tr>
<tr>
<td>Grewal et al [75], 2015</td>
<td>Postural stability (FES-1 and postural stability)</td>
<td>• PROs:</td>
</tr>
<tr>
<td></td>
<td>Diabetes peripheral neuropathy (VPT)</td>
<td>• Significant improvement in mental health component of SF-12 in intervention group but not in control group ($P$ value for intergroup differences = .04)</td>
</tr>
<tr>
<td></td>
<td>Daily PA (time spent sitting, standing, and walking and total step count) measured via shirt-worn sensor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality of life (SF-12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No significant improvement of HbA1c ($P$ value for intergroup differences = .42), BMI ($P$ value for intergroup differences = .19), lipid ratio ($P$ value for intergroup differences = .90), and mean arterial pressure ($P$ value for intergroup differences = .83) in any study arm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction in average center of mass sway area (in degrees) in intervention group but not in the control group ($P$ value for intergroup differences = .04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction in mediolateral center of mass sway area in intervention group but not in the control group ($P$ value for intergroup differences = .008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction of hip and ankle sway in the intervention group but not in the control group ($P$ value for intergroup differences = .008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant reduction of ankle sway degree in intervention group but not in control group (with and without blindfold; $P$ value for intergroup differences = .02)</td>
</tr>
</tbody>
</table>

Study quality according to RoB 2

- Low
Study quality according to RoB 2°

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Outcome (outcome measure)</th>
<th>Results</th>
<th>Nonsignificant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Höchsmann et al [28,29], 2019</td>
<td>Behavioral outcomes:</td>
<td>- No significant differences in time spent sitting ($P=.62$), standing ($P=.36$), or walking ($P=.08$) in intervention and control group</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PROs:</td>
<td>- No significant differences in physical component of SF-12 in intervention and control group ($P$ value for intergroup differences=.64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clinical outcomes:</td>
<td>- No significant differences in FES-I in intervention and control group ($P$ value for intergroup differences=.31)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No effects on VPT reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Only descriptive reporting of PA adherence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clinical parameters:</td>
<td>- No changes in resting heart rate ($P=.10$), SBP ($P=.38$), and DBP ($P=.18$) in any of the study groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No changes in skeletal muscle mass in any of the study groups ($P=.71$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No changes in total cholesterol ($P=.55$), HDL-C ($P=.46$), LDL-C ($P=.74$), and triglycerides ($P=.95$) in any of the study groups</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Statistical significant
- Nonsignificant
- Study quality according to RoB 2°
  - Low [28]
  - Moderate [29]
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Outcome (outcome measure)</th>
<th>Results</th>
<th>Study quality according to RoB 2^*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Statistically significant</td>
<td>Nonsignificant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Behavioral outcomes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant increase in intrinsic PA motivation in the intervention group ($P&lt;.001$) and nonsignificant decline in the control group ($P&gt;.05$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant between-group differences in intrinsic PA motivation in the intervention ($P&lt;.05$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant between-group differences in intrinsic PA motivation after intervention on the subscales for interest or enjoyment ($P&lt;.05$) and perceived competence ($P&lt;.05$) (significant increase in intervention group ($P&lt;.001$))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant increase in intrinsic PA motivation after intervention on the subscales for perceived choice in the intervention group ($P&lt;.05$) but no between-group differences ($P&gt;.05$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant relationship between in-game exercise (measured in minutes) and changes in intrinsic PA motivation (total score) ($P=.01$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Daily PA increase in the intervention and control group with a significant difference between groups ($P&lt;.001$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Only descriptive reporting of PA adherence</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinical parameters:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significantly higher increase in step count in intervention group than in control group ($P&lt;.001$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No changes in HbA1c in the intervention group but increase in the control group (significant between-group differences) ($P=.02$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significant increase in aerobic capacity in the intervention group and decrease in the control group (significant between-group differences) ($P&lt;.001$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Significantly higher decrease in body fat mass in intervention than in control group ($P=.045$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clinical parameters:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No significant differences in DBP, SBP, total cholesterol, HDL-C, LDL-C, triglycerides, metformin or DPP-4 inhibitor treatment, or physical well-being in any of the study groups ($P$ value not reported)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Low</td>
<td></td>
</tr>
</tbody>
</table>
Study year | Outcome (outcome measure) | Results | Study quality according to RoB 2
--- | --- | --- | ---

| | HbA₁c | Behavioral outcomes: | Low |
| | BMI | | |
| | FPG | | |
| | Total cholesterol | | |
| | LDL-C | | |
| | HDL-C | | |
| | Triglycerides | | |
| | Self-reported PA | | |
| | SBP | | |
| | DBP | | |
| | Self-assessed diabetes-related impairment using PAID | | |
| | Self-assessed physical and mental well-being using SF-12 | | |
| | Subjective well-being using WHO-5 | | |
| | Quality of life using ADS-L | | |

- **Statistically significant**
  - Significant increase in PA in both groups (P value for intergroup differences <.001)
  - Significant decrease in diabetes-related impairment in both groups (P value for intergroup differences=.03)
  - Significant improvement in subjective well-being in the intervention group but not in the control group (P value for intergroup differences=.004)
  - Significant improvement in mental health in the intervention group but not in the control group (significant between-group effects; P value for intergroup differences=.02)
  - Significant improvement in quality of life in the intervention group but not in the control group (significant between-group effects; P value for intergroup differences <.001)

- **Nonsignificant**
  - No significant differences in PPR in any group (P value not reported)
  - No significant differences in diabetes-related impairment during intervention time in any study group (P value not reported)
  - No significant differences in diabetes-related impairment and empowerment at follow-up in any study group (P value not reported)
  - No significant differences in urine microalbumin to creatinine ratio in any study group (P value not reported)

Kerfoot et al [76], 2017

- HbA₁c
- Oral diabetes medication PPR
- Urine microalbumin to creatinine ratio
- Diabetes Empowerment Scale-Short Form
- Self-assessed diabetes-related impairment using PAID

- **PROs:**
  - Significant increase in empowerment in the intervention group but decrease in the control group during intervention time (significant between-group differences; P value for intergroup differences=.01)

- **Clinical parameters:**
  - Significantly higher reduction in HbA₁c in the intervention group than in the control group both after the intervention and at follow-up (significant between-group differences; P value for intergroup differences=.048)
  - Significantly higher reduction in HbA₁c in the intervention group for patients with baseline HbA₁c of >75 mmol/mol (P=.03)
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Outcome (outcome measure)</th>
<th>Results</th>
<th>Study quality according to RoB 2&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koohmareh et al [79], 2020</td>
<td>FPG Knowledge of diabetes diet (self-developed Amoo&lt;sup&gt;c&lt;/sup&gt; test) Attention to food glucose levels and food calories</td>
<td>• Behavioral outcomes: Significantly more attention to food glucose levels and food calories in intervention group but not in the control group (P value for intergroup differences=.001)</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PROs: Significantly higher Amoo test scores in the intervention group but not in the control group (P value for intergroup differences=.001)</td>
</tr>
<tr>
<td>Maharaj et al [80], 2021</td>
<td>Self-care behaviors (SD-SCA&lt;sup&gt;d&lt;/sup&gt;) Illness beliefs (Brief Illness Perception Questionnaire)</td>
<td>• Behavioral outcomes: Borderline significant median difference in high-fat food consumption (lower in mySugr group; P value for intergroup differences=.052)</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Behavioral outcomes: No significant median differences in self-care behaviors in any study group (P value for intergroup differences=.64) PROs: No significant median differences in illness beliefs in any study group (P value for intergroup differences=.05)</td>
</tr>
<tr>
<td>Turnin et al [77], 2021</td>
<td>HbA&lt;sub&gt;1c&lt;/sub&gt; BMI Waist circumference</td>
<td>• Clinical outcomes: Significant reduction in waist circumference in intervention group but not in control group (P value for intergroup differences=.04) Stronger intergroup differences in frequent users (P=.008)</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clinical outcomes: No significant differences in HbA&lt;sub&gt;1c&lt;/sub&gt; between intervention and control group after adjustment for risk factors (P value for intergroup differences=.12) No significant differences in BMI between intervention and control group after adjustment for risk factors (P value for intergroup differences=.08)</td>
</tr>
</tbody>
</table>

<sup>a</sup>RoB 2: Cochrane risk-of-bias tool version 2.
<sup>b</sup>BDNF: brain-derived neurotrophic factor.
<sup>c</sup>VEGF: vascular endothelial growth factor.
<sup>d</sup>IGF-1: insulin-like growth factor–1.
<sup>e</sup>HbA<sub>1c</sub>: glycated hemoglobin (blood sugar value).
<sup>f</sup>PA: physical activity.
<sup>g</sup>BMI: body mass index
<sup>h</sup>PRO: patient-reported outcome.
<sup>i</sup>FES-I: Falls Efficacy Scale–International.
<sup>j</sup>VPT: vibration perception threshold.
<sup>k</sup>SF-12: Short Form Health Survey.
<sup>l</sup>LDL-C: low-density lipoprotein cholesterol.
<sup>m</sup>HDL-C: high-density lipoprotein cholesterol.
<sup>n</sup>SBP: systolic blood pressure.
<sup>o</sup>DBP: diastolic blood pressure.
<sup>p</sup>FPG: fasting plasma glucose.
<sup>q</sup>PAID: Problem Areas in Diabetes Scale.
<sup>r</sup>WHO-5: 5-item World Health Organization Well-Being Index.
<sup>s</sup>ADS-L: Allgemeine Depressionsskala long version (German).
<sup>t</sup>DPP-4: dipeptidyl peptidase 4.
Effects of Gamified Health Interventions on PROs

In the included RCTs, data were provided on health literacy in general [74,79,80], quality of life [75,78], diabetes-related impairment [76,78], and subjective well-being [78].

Concerning health literacy, Koohmareh et al [79] found a significant positive effect of their educational game on knowledge concerning a diet adequate for patients with diabetes (measured on a self-developed scale; $P=.001$). Maharaj et al [80] found no significant median differences in illness beliefs (measured using the Brief Illness Perception Questionnaire) between the gamified mySugr and the self-management app without any game components ($P$ value for intergroup differences=.05). Glasgow et al [74] used health literacy as a moderating variable in a multivariate analysis of covariance to measure the effects of their self-management website with quiz elements but found no interaction effects ($P$ value not reported).

In the domain of quality of life, Grewal et al [75] found a significant improvement in mental well-being because of the virtual balance training they studied ($P$ value for intergroup differences=.04) but no effect of said intervention on physical well-being ($P$ value for intergroup differences=.64; both measured using the Short Form 12 Health Survey). The same effect was demonstrated by Kempf and Martin [78], who studied Wii Fit Plus games, using both the Short Form 12 Health Survey and a German depression scale. Subjective well-being, measured using the 5-item World Health Organization Well-Being Index scale, improved significantly in the study by Kempf and Martin [78] as well ($P$ value for intergroup differences=.004).

Grewal et al [75] and Kempf and Martin [78] studied the effects of gamified interventions on diabetes-related impairment both using the Problem Areas in Diabetes Scale. Although the latter found a positive effect of the Wii Fit Plus games ($P$ value for intergroup differences=.03) [78], the former found no effect of virtual balance training ($P$ value not reported) [75]. In terms of impairment, Grewal et al [75] found no effect of the virtual balance training on fear of falling ($P$ value for intergroup differences=.31).

Overall, sparse positive effects of digital health interventions with game components could be found on health literacy and diabetes-related impairment, whereas substantial evidence was found for the improvement of subjective mental well-being.

Effects of Gamified Health Interventions on Clinical Outcomes

The clinical outcome studied most often in the included RCTs was $HbA_1c$ [29,73,74,76-78]. Furthermore, 30% (3/10) of the included studies analyzed changes in BMI [74,77,78], whereas Höchsmann et al [28], Glasgow et al [74], and Kempf and Martin [78] also analyzed lipid outcomes and blood pressure values. Fasting plasma glucose (FPG) was studied twice [78,79], and so was the heart rate of the participants [28,72]. Brinkmann et al [72] studied cognitive parameters in conjunction with lactate values, and Grewal et al [75] analyzed postural stability.

The effects of the gamified interventions on $HbA_1c$ levels studied in 60% (6/10) of the RCTs are depicted in Figure 3, where asterisks mark significant changes in $HbA_1c$. Only Kerfoot et al [76] ($P$ value for intergroup differences=.048) and Kempf and Martin [78] ($P$ value for intergroup differences<.001) found significant positive effects of a mobile, team-based DSME quiz and Wii Fit Plus games, respectively, on $HbA_1c$ levels in terms of a reduction in the intervention group and significant between-group effects. Dugas et al [73] found a significant effect of time spent using the self-management app with an award system on $HbA_1c$ levels in a regression analysis ($P<.01$), as well as an interaction effect of total adherence score (additive score of exercise, nutrition, and medication adherence; $P<.01$).
Figure 3. Effects of the interventions studied on glycated hemoglobin (HbA1c) levels. *Significant. Δ HbA1c: difference in HbA1c between pre- and postintervention measurement; CG: control group; DSME: Diabetes Self-Management and Education; IG: intervention group; n.s.: not significant [29,73,74,76-78].

Kempf and Martin [78] also found significant positive effects of the intervention on BMI (P value for intergroup differences=.008); however, Glasgow et al [74] (P value for intergroup differences=.20) and Turnin et al [77] (P value for intergroup differences=.08), also studying DSME software with quiz components, did not. Significant effects on FPG after using the same software were also reported by Kempf and Martin [78] (P value for intergroup differences=.008) but not by Koohmareh et al [79] (P value for intergroup differences=.63), who studied the same intervention type. Significant effects were not found on lipid outcomes and blood pressure values [29,74,78] or on cognitive outcomes and lactate values [72]. Grewal et al [75] found improved postural stability to be an effect of the virtual balance training they studied (P=.009).

Turnin et al [77] found a significant reduction in waist circumference in the intervention group when compared with the control group both before and after adjusting for confounding factors such as baseline HbA1c levels, age, sex, or obesity (P value for intergroup differences=.04). The authors also reported an increased effect on waist circumference in those with a higher frequency of use (P=.008).

No significant effects were found on heart rate [29,72], blood pressure (systolic or diastolic) [29,78], or lipid values (i.e., on total cholesterol, triglycerides, low-density lipoprotein cholesterol, or high-density lipoprotein cholesterol) [29,78] in any of the RCTs that reported on these values.

Brinkmann et al [72] studied the vascular endothelial growth factor as an indicator of diabetes retinopathy but found no effect of the Wii Fit exergame they studied (P value not reported).

Brinkmann et al [72] found improved lactate values in the intervention group (P=.04) but not in the control group (P value not reported), which used a stationary bicycle, thus indicating higher fitness levels owing to the exergame. Improved fitness because of an exergame was also demonstrated by Höchsmann et al [29] using aerobic capacity as an indicator (P<.001). Grewal et al [75] found improved postural stability to be an effect of the virtual balance training they studied (P=.009).

As Brinkmann et al [72] aimed to also improve cognitive functioning of the study participants by adding cognitive challenges to the exergame, they studied the brain-derived neurotrophic factor as an indicator of learning and memory capacity and found no significant effect of the intervention (P value not reported).

Overall, sparse positive effects of digital health interventions with game components could be found on clinical outcomes, with fitness-related outcomes improving the most.

A complete overview of the effects found in the included RCTs can be found in Table 3. If not stated otherwise, differences in outcomes are reported between the intervention and control group(s) before and after the study in Table 3.

Discussion

Principal Findings

Despite heterogeneity in the effects of the gamified applications studied in this review on behavioral, clinical, and patient-reported outcomes, certain patterns emerged. Exergames had the tendency to improve fitness-related and, to a smaller
extent, clinical values (as shown in 4/10, 40% of the RCTs [73, 76-78]), whereas educational games affected disease-related knowledge and especially nutrition behavior (as shown in 4/10, 40% of the RCTs [73, 74, 79, 80]). In addition, exergames had a potential to improve outcomes related to self-reported well-being such as quality of life and diabetes-related impairment (as shown in 2/10, 20% of the RCTs [75, 78]).

Comparison With Prior Work

The results confirm those of a meta-analysis by DeSmet et al [82], who showed significant positive effects of serious games on health behavior independent of any diagnosis and its theory-based determinants and significant but much smaller effects on various clinical outcomes. As for nutrition behavior, which improved according to 40% (4/10) of the RCTs included in this review solely because of self-management applications with quiz elements, Ledoux et al [83] found similar effects of a serious game for young patients with type 1 diabetes.

Somewhat surprisingly, self-efficacy was not among the outcomes studied in the included RCTs despite existing evidence that both exergames and serious games can increase the feeling of being able to actively affect one's health outcomes [65, 84]. A precondition derived from behavior change theories is the option for goal setting by the intervention participants themselves as opposed to behavioral goals predefined by the intervention developers [85]. This precondition was met by all the gamified interventions studied in this review that used a reward or scoring system as these game components allow for the autonomous setting of target values that the users aim to achieve and, thus, generate intrinsic motivation [12].

A coherent narrative as an instrument of storytelling was used in only 20% (2/10) of the included interventions; however, it helped in producing positive effects on behavioral, knowledge-related [28, 79], and clinical outcomes [29]. As such, the results confirm observations made for narrative health communication messages when delivered in a digital manner [86].

Physical fitness because of an increase in PA owing to interventions with game components is an especially promising result as PA is one of the primary target behaviors of diabetes self-management. The same is true for dietary behavior [6]. As such, these results underline once more the potential of digital interventions for DSME, acknowledged also by the ADA and the European Association for the Study of Diabetes in a joint statement [87].

The relatively high mean age of the study participants is surprising insofar as the traditional target group for digital health applications of all sorts is usually younger people [88]. Christensen et al [17], in sensitivity analyses, also found a positive effect of game-based interventions only for people aged <18 years. Especially regarding exergaming, the positive effects for older people because of high engagement have already been proven elsewhere [89]. Given the fact that most of the gamified interventions studied in this review (7/10, 70%) were designed for mobile devices, the high penetration of such devices in all age groups might play an important role in overcoming the digital health divide because of age [90]. Therefore, the results need to be considered in light of demographic changes and a rise in the demand for health care, especially for chronic diseases [91].

With intervention duration varying widely (Table 2), statements on its relevance have to be made with caution. However, the results show that effects on clinical outcomes can be achieved via digital health interventions with game components after at least 12 months [76, 78], which corresponds to the concept of DSME as a continuous, long-term intervention [4, 6]. However, even 36 weeks of intervention duration did not guarantee significant effects on BMI and HbA1c [74], hinting at a washout in intervention fidelity common in digital health applications [92].

The fact that generally positive effects of digital health interventions of any kind are biased because of low overall study quality also mirrors findings of recent evidence syntheses [22, 93]. The methodological issues concerning the quality of obtainable evidence raised in the joint statement by the ADA and the European Association for the Study of Diabetes [87] persist in this review even though all the included studies (10/10, 100%) were RCTs or had adaptive RCT designs. The issue of blinding of study participants and personnel to the allocation is common in digital health trials as the fact that one did not receive a gamified health application is easily uncovered [8]. Therefore, the study conducted by Maharaj et al [80] is a positive example of comparing 2 digital health applications whereby one is augmented with game components. The issue of missing outcome data because of dropouts, well known to researchers in multiple fields, is also common in digital health research [94] and is linked to the issue of intervention fidelity. The reasons as to why trial participants may lose interest in the use of digital health applications can be numerous [92]. Therefore, participatory design is an important precondition for gamified applications as well [95]. Nonvalidated outcome measures, another major source of bias in the included RCTs, point to the need to develop core outcome sets for digital health applications with or without game components.

Strengths and Limitations

Robust and reproducible systematic review methods were used to identify the best available evidence, and the results were reported according to the PRISMA checklist (Multimedia Appendix 3 [23]). Owing to the fact that all steps taken in this review were performed by 2 researchers independently and checked by a third researcher, it is highly unlikely that relevant records were overlooked or incorrectly discarded as irrelevant. The same is true for information in the included records relevant to the quality assessment.

The focus on research published in German or English is a limitation. Owing to the diverse intervention types (with intervention durations varying widely) and outcomes studied in the rather large number of included RCTs (compared, eg, with the meta-analysis by Christensen et al [17]), a meta-analysis especially of clinical outcomes was not deemed feasible. Rather, the broad realm of outcomes studied allows for a holistic overview of the potential effects of digital health interventions with game components.
Some of the included RCTs (3/10, 30%) had a considerable loss to follow-up, which would have required intention-to-treat analyses. However, most studies (7/10, 70%) conducted per-protocol analyses instead. Furthermore, the reasons for dropout were not always reported.

Little to no follow-up time after the intervention period limits the comparability of the gamified interventions studied in the included RCTs with other digital or analog behavior change interventions that have demonstrated sustainable effects over time [96]. However, according to both theory and evidence, sustainable and long-lasting behavior change over a time span of at least 6 months is necessary to achieve improved health outcomes and still not easily achieved [85]. Owing to limited reporting of intervention intensity, no discernible patterns regarding effectiveness could be found. Moreover, effective components of gamified health interventions regarding both clinical outcomes and behavior change have not been identified yet [97].

**Conclusions**

This systematic review provided a thorough analysis of the effectiveness of digital health interventions with game components for the self-management of type 2 diabetes. The included RCTs analyzing exergames showed positive effects on fitness-related outcomes and, albeit only in 1 case, also on HbA1c. Educational games improved dietary habits and subjective mental health and well-being. However, the evidence base was ambiguous and further limited because of the considerable risk of bias in the study designs of most of the included RCTs (9/10, 90%). Nevertheless, the results imply that digital health interventions with game components can help improve PA, dietary habits, and well-being. Therefore, these applications, when developed based on theory and evaluated rigorously, can help achieve the behavioral goals mentioned in several guidelines for DSME. Given the mostly low quality of the included RCTs, the presented evidence allows for nothing more than the suggestion of digital health interventions with game components as a supplement to traditional DSME, which corresponds to the statement by Fleming et al [87]. The main results of this review are summarized in **Textbox 3**.

**Textbox 3. Main results of the review.**

- The variety of game components used in digital health interventions for type 2 diabetes ranges from quiz components over storytelling elements to full-scale exergames.
- Digital health interventions containing game components might improve the health behavior patterns, quality of life, and clinical outcomes of patients with type 2 diabetes.
- The methodological quality was low in most of the included randomized controlled trials.

Further research should compare digital health interventions that contain game components with those that do not rather than comparing the former with usual care and, as such, risking bias because of high dropout rates. Furthermore, longer follow-up assessments are crucial to detect whether the effects of gamified interventions are sustainable. Adaptive study designs such as microrandomization, where individuals are randomized to several intervention components or treatments for the study duration [98], could be used to determine which game components have an effect on which outcome domain (behavioral, PROs, or clinical). Larger sample sizes would allow for more detailed subgroup analyses [99] to determine which user groups, in addition to those defined by age, profit the most from digital health interventions containing game components. Finally, but importantly, a comparison of the effects of digital health interventions with game components for patients with type 1 and 2 diabetes is bound to provide further insights.

**Acknowledgments**

The work for this review was partly funded by the county of Saxony, Germany.

**Authors’ Contributions**

LO and LH devised the study design and conducted the literature search, study selection process, quality assessment, and data extraction together with TH. SD and PT were involved in the consensus process during study selection, data extraction, and quality assessment. PT, OS, and MS analyzed the extracted data and derived points for discussion and implications. LH drafted the manuscript, which was critically revised for important intellectual content and approved by all the authors.

**Conflicts of Interest**

Unrelated to this study, OS served as a consultant for Novartis. Unrelated to this study, JS reports institutional grants for investigator-initiated research from the German Federal Joint Committee, Federal Ministry of Health, Federal Ministry of Education and Research, European Union, Federal State of Saxony, Novartis, Sanofi, Allergologisk Laboratorium København (ALK), and Pfizer. He also participated in advisory board meetings as a paid consultant for Sanofi, Lilly, and ALK. JS serves the German Ministry of Health as a member of the Sachverständigenrat Gesundheit und Pflege (expert advisory board on health and care).
References


Abbreviations

ADA: American Diabetes Association  
CASP: Critical Appraisal Skills Programme  
DSME: Diabetes Self-Management and Education  
FPG: fasting plasma glucose  
HbA1c: glycated hemoglobin  
PA: physical activity  
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
PRO: patient-reported outcome  
RCT: randomized controlled trial  
RoB 2: Cochrane risk-of-bias tool version 2

©Linda Ossenbrink, Tina Haase, Patrick Timpel, Olaf Schoffer, Madlen Scheibe, Jochen Schmitt, Stefanie Deckert, Lorenz Harst.  
Originally published in JMIR Serious Games (https://games.jmir.org), 01.06.2023. 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.
Comparison of Exergames Versus Conventional Exercises on the Health Benefits of Older Adults: Systematic Review With Meta-Analysis of Randomized Controlled Trials

Xi Chen¹*, MD; Lina Wu¹*, MD; Hui Feng¹,²,³, Prof Dr, PhD; Hongting Ning¹, PhD; Shuang Wu¹, PhD; Mingyue Hu¹, PhD; Dian Jiang¹, BSc; Yifei Chen¹, MD; Yu Jiang⁴, BSc; Xin Liu⁵, MD

¹Xiangya School of Nursing, Central South University, Changsha, China
²Xiangya-Oceanwide Health Management Research Institute, Central South University, Changsha, China
³National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Changsha, China
⁴Changsha Xingsha Hospital, Changsha, China
⁵Department of General Practice, 921 Hospital of Joint Logistics Support Force, The Chinese People's Liberation Army, Changsha, China

* these authors contributed equally

Corresponding Author:
Xin Liu, MD
Department of General Practice
921 Hospital of Joint Logistics Support Force
The Chinese People's Liberation Army
No.1 Hongshan Bridge
Changsha, 410008
China
Phone: 86 13308490423
Email: 281039313@qq.com

Abstract

Background: Conventional exercises (CEs) can provide health benefits for older adults, but the long-term exercise adherence rate is low. As an emerging, stimulating, and self-motivating strategy, exergames (EGs) are defined as combinations of exercises and games that users carry out through physical actions. They can promote exercise, but the health effects of EGs versus CEs on the physical function and mental health (cognitive function, depression, and quality of life) of older adults remain controversial.

Objective: The aim of the study is to compare the health benefits of EGs versus those of CEs for the physical function and mental health of older adults.

Methods: A comprehensive search was conducted from the earliest available date to February 2023 in the following 6 databases: PubMed, Web of Science, Embase, Cochrane, CINAHL, and PsycINFO. All English-language randomized controlled trials comparing the effects of EGs versus those of CEs on the physical function and mental health of older adults, with nearly same physical activity between the 2 interventions, were included. Risk of bias was independently evaluated by 2 authors using the Cochrane risk of bias in randomized trials tool. Two authors independently extracted data. We followed the Cochrane Handbook of Systematic Reviews of Interventions to process and analyze the data for meta-analysis. Standardized mean differences (SMDs) and 95% CIs were used for continuous data, and random models were used for analyses.

Results: We included 12 studies consisting of 919 participants in total. Of these, 10 studies were eventually included in the meta-analysis. The results showed that EGs versus CEs exhibited no significant differences in physical function (P=.13; τ²=0.31; χ²=26.6; I²=77%; SMD=0.37; 95% CI –0.11 to 0.86) or cognitive function (P=.63; τ²=0.01; χ²=3.1; I²=4%; SMD=0.09; 95% CI –0.27 to 0.44) effects.

Conclusions: Our findings indicate no significant difference between EGs and CEs in improving the physical function and cognitive function of older adults. Future studies are required to compare the effects of EGs versus those of CEs on cognitive function according to cognitive status, quantify the “dose-effect” relationship between EGs and health benefits, and evaluate the effects of different types and devices of EGs with regard to the health benefits of older adults.

Trial Registration: PROSPERO CRD42022322734; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=322734
**Introduction**

**Background**

With fertility rates declining and life expectancies rising, the global population is aging [1]. The number of adults older than 65 years has tripled over the past 50 years and, by 2050, older adults will make up a quarter of the global population [2-4]. Aging can lead to degenerative changes in the physical and cognitive function of older adults, resulting in impaired daily life functions and reducing the independence of older adults, thereby affecting their mental health and increasing the burden of health care [5]. Physical dysfunction is increasingly common at end of life [6], and the global pooled incidence rate of older adult frailty in communities is 43.4 per 1000 person-years [7]. Mental health problems are also common among older adults in China, and it is reported that 21.129 out of 88,417 older adults (23.6%) have these problems [8].

The physical and mental decline of older adults may eventually have serious social and economic consequences for an aging society [9]. For physical decline, frail older people (97/177, 54.5%) needs more health care services than nonfrail older people (30/987, 2.2%) [10,11], and the median hospitalization cost of frail patients is more than twice that of healthy patients (US $44,408 vs US $18,660) [12]. Furthermore, frail people also require continuous care at 5.82-fold the rate of healthy people after discharge [13]. Regarding mental health, studies predict that worldwide care costs for dementia, which is associated with cognitive decline, will increase to US $2 trillion by 2030 [14]. At the same time, the global economic burden of mental disorders in 2010 was similar to that of cardiovascular diseases, higher than that of cancer, and is expected to more than double by 2030 (US $2.5 trillion vs US $6.1 trillion) [15]. Thus, measures must be taken to promote healthy aging. A proven effective strategy is regular physical exercise [16].

Exercise is defined as “planned, organized and repetitive physical activity” [17]. Several studies have shown that conventional exercises (CEs), such as aerobic, resistance, and combined exercise, can improve the physical function and mental health of older adults [18-23]. However, older adults may not exercise due to lack of access to sports venues (eg, the COVID-19 pandemic), inconvenience, or lack of motivation [24]. In addition, the long-term exercise adherence rate for CEs among older adults seems to be low [25,26]. Therefore, an increasing number of studies have considered possible alternatives to CEs.

Exergames (EGs) are a combination of exercise and gaming that allows people to physically interact with a web-based game scene on a screen [27]. It requires the player’s physical performance, as the technology used in the game system tracks the player’s movements to control those in the game, thus immersing the player [28]. For example, in a Kinect-based EG [29], the game uses infrared light and cameras in the Kinect system to capture and track the player’s movements and creates a full-body 3D web-based map, which is rendered by the screen in front of the player. Participants stand in front of the screen, imitate the actions of digital characters on the screen, and adjust their movements in real time based on instant visual and auditory feedback. It can be implemented in community centers, retirement institutions, long-term care facilities, assisted living, nursing homes, burn centers, hospitals, and homes [30,31]. EGs are an interesting strategy for active aging and good mental health [32]. They have been proven to be acceptable, feasible, safe, enjoyable, stimulating, and self-motivating tools [30,33-35]. EGs can be carried out at home, alone, or in groups, which may make it easier for older adults to participate in exercise [35]. In terms of physical function and mental health, in some studies, it is found that the EGs are better than CEs [36-38]; some have reported that EGs are as effective as CEs [39,40]; yet others have concluded that the effects of CEs are better [41]. In short, there is controversy regarding the effects of EGs and CEs on the physical function and mental health of older adults.

**Research Gap and Aim**

While 2 systematic reviews have compared the impacts of EGs and CEs on older adults, the results were inconclusive [42,43]. One compared balance and prevention of falls for EGs versus CEs in healthy older adults, and 20 randomized controlled trials (RCTs) were included [42]. EGs were found to have greater improvements in posture control and dynamic balance than CEs [42]. The other review compared the impact of EGs versus CEs on the cognitive function of older adults; it included 13 studies for systematic review and 11 studies for meta-analysis [43]. The results showed no statistical difference between the EGs and CEs in cognitive function [43]. However, these systematic reviews (1) did not distinguish sedentary video games from EGs; (2) searched only PubMed, Web of Science, and Cochrane databases; (3) included non-RCTs and quasi-experiments; (4) did not consider inconsistencies in exercise content between EGs and CEs; and (5) failed to distinguish the effects of EGs alone from the effects of EGs combined with other interventions, such as CEs.

Therefore, a systematic review should be conducted examining all the evidence, using quantitative analysis to compare the impacts of EGs versus CEs on older adults. Our study aims to compare the health benefits of EG versus CE programs for older adults’ physical function and mental health (cognitive function, depression, and quality of life [QOL]). The comprehensive research results may provide a basis for the choice of rehabilitation strategy for the healthy aging of older adults.
Methods

The PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) was used to report this review [44].

Search Strategy

PubMed, Web of Science, Embase, Cochrane, CINAHL, and PsycINFO were searched using subject headings and keywords from their inception up to February 2023. We limited the publication language to English. Search strategies are shown in Multimedia Appendix 1. The references of all included studies were reviewed for further relevant research. If more information about relevant research was needed, we contacted the first author.

Criteria for This Review

Types of Studies
Published peer-reviewed reports of RCTs were included. We considered trials in which randomization was implied with at least 2 intervention arms (ie, EGs and CEs). Quasi-randomized studies were excluded. Abstracts, systematic reviews, case reports, and registered trial reports were also excluded.

Participants

Studies focused on older adults, where all participants aged 60 years or older, were included. Studies with a hybrid sample (ie, younger and older adults) and older adults with hemiplegia or other paralysis were excluded.

Types of Interventions

Activities carried out under sitting conditions and controlled by handheld devices (ie, sedentary video games) were excluded. There must be a group where the only intervention is EGs not combined with any other intervention, such as CEs or cognitive training.

Studies comparing EGs with CEs (eg, aerobic and endurance training, resistance or strength training, multicomponent training, balance training, high-intensity interval training, Tai Chi, yoga, dance, Otago, physical therapy exercises, ball exercise, and treadmill) were included. The CEs performed precisely the same physical activity as the EGs but did not involve web-based feedback.

Types of Outcomes

Primary outcomes
Physical function is defined as the ability to perform and complete objectively measured performance-based tasks that assess cardiovascular fitness, muscle strength, flexibility, mobility, and balance [45]. Physical function was measured by the gait speed test, Berg balance test, sit-to-stand test, or 30-second chair stand test.

Secondary Outcomes

Cognitive function is defined as a broad set of thinking abilities that can be measured using performance-based tasks [23]. Cognitive function was measured by the Montreal Cognitive Assessment (MoCA) or Mini-Mental Status Exam (MMSE). If a study used both MoCA and MMSE to measure cognitive function, we extracted only the data measured by MoCA because a previous study has shown that MoCA is a better cognitive function measurement method and can detect cognitive heterogeneity well [46].

Depression was measured by the Geriatric Depression Scale. QOL was measured by SF-36 (36 health survey questionnaire).

Screening Process and Data Extraction

Screening Process

Two reviewers (XC and LW) independently conducted literature screening, first screening titles and abstracts to determine whether the research met the inclusion criteria.

Then the full text was obtained to determine whether the study was eligible for inclusion. If 2 reviewers had disagreements, a third reviewer was to be consulted to decide whether to include it. If disagreements could not be resolved through discussion, we would attempt to contact the corresponding author of the study for clarification.

Data Extraction

Two authors extracted data independently. The extracted data included first author, age, sample size (female %), population type, dosage of intervention, types of intervention, types of control, outcome, device, and results for the review objectives. We extracted data presented in figure and graph form when 2 review authors independently obtained the same results. Disagreements were resolved by a third reviewer. If data were missing, we would contact authors.

Quality Assessment

Methodological quality was assessed independently by 2 authors using the Cochrane risk-of-bias tool [47], which includes the following seven contents items: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) incomplete outcome data, (5) selective reporting, and (6) other bias.

Statistical Analysis

We followed the Cochrane Handbook of Systematic Reviews of Interventions to process and analyze data for meta-analysis [47]. The results of EGs and CEs were compared after the intervention. All results are continuous variables. Meta-analysis was performed using Review Manager 5.4 software (Cochrane). Standard mean difference and 95% CIs were used for continuous data. Subgroup analyses based on gender distribution were performed. The results were regarded as statistically significant when \( P < 0.05 \) [47]. The heterogeneity test was quantified using the \( I^2 \) statistic and the chi-square \( P \) value. The \( I^2 \) statistic was considered low, moderate, or large at 25%, 50%, or 75%, respectively [48]. A chi-square \( P \) value of .05 or less suggests heterogeneous meta-analyzed studies [47]. The random model was selected because the included studies from different populations were heterogeneous. Publication bias was assessed by examining funnel plots.
Results

Selected Studies
In total, 6410 potentially relevant studies were identified through database searching. After deduplication, the titles and abstracts of 4089 records were screened by checking the inclusion criteria; 256 studies were further screened by viewing the full texts for eligibility; 12 papers were included for the systematic review, and 10 papers were included for meta-analysis. Figure 1 shows the PRISMA flow diagram for paper selection.

Five studies focused on community-dwelling older adults [38,49-52], 3 studies recruited older adults from care facilities [29,40,53], 1 study was conducted in hospital and home [54], 1 study recruited older adults in a clinical physical rehabilitation unit [55], and 2 studies did not report the research setting [56,57]. Multimedia Appendix 2 [29,38,40,49-57] presents the characteristics of the included studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flowchart of the study selection process. RCT: randomized controlled trial.
Characteristics of Participants and Interventions

The 12 included studies had a total sample size of 919 participants, with individual studies ranging from 18 to 282 participants. One study only recruited women [56], 1 study had an equal proportion of female and male participants [55], 6 studies recruited more female participants [29,40,49,52,53,56], and 5 studies recruited more male participants [38,50,51,54,57]. In total, 6 studies focused on older adults with Parkinson disease [38,50,51,54,55,57], 1 on older adults with mild cognitive impairment [49], and 1 on frailty [29]; 4 investigated older adults without special conditions [40,52,53,56].

A 2-arm design was used in 11 studies, including an intervention arm and a control arm [29,38,40,50-57]; 1 study used a 3-arm design, including 2 intervention arms and a control arm [49]. The intervention duration ranged from 5 to 12 weeks, and the most widely used duration was 6 weeks. The frequency was 2 or 3 times per week, and the length of each session was 30 to 90 minutes. The older people in 2 studies only participated in a single session of exercise training [52,53]. In total, 4 studies used Nintendo Wii for the intervention [53,54,56,57], 5 studies used Kinect for the intervention [29,38,49-51], 1 study used Tymo for the intervention [55], 1 study used GRAIL for the intervention [52], and 1 study did not report the intervention device [40].

Meta-Analysis Results

Due to the lack of original data, we did not perform a meta-analysis of 2 papers [38,50].

**Primary Outcome**

Seven studies reported physical function according to the 30-second chair stand test, gait assessment (GaitRite, CIR Systems, United States), the sit-to-stand test, or the Berg balance test [40,51,52,54,57]. There was no significant difference between EGs and CEs ($P=0.13$; $\chi^2=0.31$; $\chi^2=26.6$; $I^2=77%$; standardized mean difference [SMD]=0.37; 95% CI –0.11 to 0.86), more details in Figure 2.

**Secondary Outcome**

**Cognitive Function**

Four studies reported cognitive function according to the MoCA or MMSE [29,40,49,57]. No significant difference was observed between EGs and CEs in MoCA or MMSE ($P=0.63$; $\chi^2=0.01$; $\chi^2=3.1$; $F^2=4%$; SMD=0.09; 95% CI –0.27 to 0.44), more details in Figure 3.

**Depression**

A single study reported depression and found no significant difference between EGs and CEs [40].

Among the control group exercises, 1 study consisted of squats, postural displacements, dance, and sports (volleyball and boxing) [53]; 1 included resistance exercise, aerobic exercise, Tai Chi, and balance exercises [29]; 1 included global exercises and balance exercises [57]; 1 used Tai Chi [49]; 1 included task-oriented exercise, walking, stretching, balance training, flexibility exercises, and coordination exercises [55]; 1 included passive range of motion for lower extremities, active free exercises, stretching exercises, strength training, resisted exercise, and balance training [54]; 2 included strength exercises and core training [40,56]; and 4 studies used treadmill training in the control group [38,50].

Quality of Life

Only 1 study reported QOL [38]. The results show that the SF-36 scores of the EG group improved more than those of the CE group.

---

**Figure 2.** Meta-analysis of the effect of exergames versus conventional exercises on physical function [51-57].

**Secondary Outcome**

**Cognitive Function**

Four studies reported cognitive function according to the MoCA or MMSE [29,40,49,57]. No significant difference was observed between EGs and CEs in MoCA or MMSE ($P=0.63$; $\chi^2=0.01$; $\chi^2=3.1$; $F^2=4%$; SMD=0.09; 95% CI –0.27 to 0.44), more details in Figure 3.

**Figure 3.** Meta-analysis of the effect of exergames versus conventional exercises on cognitive function [29,49,53,57].

**Depression**

A single study reported depression and found no significant difference between EGs and CEs [40].
**Subgroup Analysis**

To further compare the effects of EGs versus CEs on physical and cognitive functions in different gender distributions, the results of subgroup analysis are shown in Figures 4 and 5. No significant difference was observed in physical function between EGs and CEs when the percentage of females <50% ($P=.23; \chi^2=0.02; \chi^2=2.4; I^2=18%; SMD=0.18; 95\% CI -0.12 to 0.49$) and when the percentage of females ≥50% ($P=.40; \chi^2=0.79; \chi^2=19.5; I^2=85%; SMD=0.41; 95\% CI -0.54 to 1.36$). No significant difference was observed in cognitive function between EGs and CEs when the percentage of females <50% ($P=.59$) and when the percentage of females ≥50% ($P=.42; \chi^2=2.3; I^2=13%; SMD=0.18; 95\% CI -0.26 to 0.62$).

**Figure 4.** Subgroup analysis of the effects of exergames versus conventional exercises on physical function [51-57].

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Study or Subgroup</th>
<th>Experimental</th>
<th>Control</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of females &lt;50%</td>
<td>Alammar et al. 2022</td>
<td>42.61 ± 9.69</td>
<td>42.63 ± 5.09</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Palomar, E 2020</td>
<td>109.9 ± 4.5</td>
<td>107.6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Pompeu, J E 2012</td>
<td>54.2 ± 2.2</td>
<td>53.1</td>
<td>4</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td></td>
<td>129</td>
<td>134</td>
<td>44.6%</td>
</tr>
<tr>
<td>Percentage of females ≥50%</td>
<td>Bobinski et al. 2022</td>
<td>18.0 ± 1.7</td>
<td>18.3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Montemayor-Juárez, R S 2015</td>
<td>3.3 ± 0.9</td>
<td>3.9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Montemayor-Juárez, R S 2017</td>
<td>4 ± 0.9</td>
<td>3.9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Zúñiga-Luna, A 2022</td>
<td>1.2 ± 1.2</td>
<td>0.8</td>
<td>30</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td></td>
<td>60</td>
<td>65</td>
<td>53.2%</td>
</tr>
</tbody>
</table>

**Figure 5.** Subgroup analysis of the effects of exergames versus conventional exercises on cognitive function [29,49,53,57].

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Study or Subgroup</th>
<th>Experimental</th>
<th>Control</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of females &lt;50%</td>
<td>Pompeu, J E 2012</td>
<td>22.2 ± 4.5</td>
<td>23.1</td>
<td>16</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td></td>
<td>16</td>
<td>16</td>
<td>25.2%</td>
</tr>
</tbody>
</table>

**Publication Bias**

The funnel plot did not show a clear funnel shape in physical function (Figure 6). The reason may be that some small studies with negative results may not favor the EG intervention.
Quality Assessment

Figure 7 shows the results of the methodological quality assessment. The quality of the included studies was found to be acceptable. Regarding the risk-of-bias assessment [47], we found that 9 studies showed a high risk of performance bias [29,38,40,50,53,57], while 2 studies had unclear bias risk [49,56]. In total, 11 studies used a single-blind protocol [29,38,40,49-55,57], and 1 used a double-blind protocol [56]. Due to the EGs and CEs, it was not possible to blind patients or study personnel to the group allocation. High risk studies may overestimate the effect of EGs compared to CEs.
Discussion

**Principal Findings**

By ensuring that EG and CE groups in the reviewed studies performed the same physical activity, this study is the first to compare the health benefits of EGs and CEs for older adults. We observed that EGs show potential as a novel approach for enhancing physical and cognitive function in older adults. The results of the meta-analysis show no significant difference in physical function or cognitive function between EGs and CEs after intervention. This means that EGs may replace CEs in these aspects. Our findings provide evidence of the beneficial effects of EGs, which may offer a promising strategy for promoting healthy aging in older adults. Given that EGs are an innovative, fun, and relatively safe form of exercise [58], this review presents timely evidence that suggests EGs could be a valuable tool for health professionals, such as physical therapists and occupational therapists.

The results show no significant difference in cognitive function between EGs and CEs. This supports the findings of previous studies [49,59,60]. One RCT with older individuals with dementia showed no significant difference in executive function, episodic memory, or working memory between EGs and aerobic training groups [60]. The other RCT suggested that the Kinect adventures training group and conventional physical therapy group had no significant difference in cognitive function after intervention, and both had positive effects on cognitive function in older adults [59]. In addition, EGs are inexpensive [61], safe [30,62], and easy to use [63], and healthy older adults living in the community can use them without supervision [64,65]. Therefore, the application of EGs in the cognitive rehabilitation of older adults should be promoted, especially for those with mild cognitive impairment or dementia.

However, our findings are inconsistent with a previously published systematic review and meta-analysis that compared the effects of EGs versus CEs on cognitive skills [43]. Based on that study, EGs seem to be more effective than conventional physical training for global cognitive performance. A possible reason is the potential ceiling effect [66]. The previous systematic review and meta-analysis included more patients with mild cognitive impairment or dementia, so our study included more older adults with normal cognitive function. Therefore, even before the intervention, the cognitive function of older adults in our study was quite good, which may make the improvement of cognitive function following EGs and CEs not obvious, resulting in a small difference in the improvement of cognitive function between the 2 groups. Thus, it is suggested that subsequent studies compare the effects of EGs and CEs on cognitive function according to the classification of the cognitive status of older adults.

Furthermore, EGs and CEs showed no significant difference in physical function. This is in accordance with previous findings [67,68]. This may be related to the impact of EGs on heart rate and energy expenditure similar to CEs [69]. Due to a lack of studies, we cannot analyze the impact of potential moderators (intervention time, intensity, and type of EG) on physical function. First, a subgroup analysis of the relationship between intervention time and effect shows that a weekly intervention affects physical function [42]. In that regard, there was no consensus on the duration of weekly interventions in the studies we included. Second, the intensity of EGs must be at least moderate to achieve the effect [69], while the included studies rarely measured the intensity of EGs. Finally, the types and devices of EGs will also affect the intervention effect. It is suggested that future studies explore the impact of different types and devices of EGs on the physical function of older adults. Subsequent studies should quantify the “dose-effect” relationship between EGs and health benefits in older adults, derive optimal intervention doses for EGs (eg, program period, weekly intervention duration, frequency, and intensity), and determine how different types and devices of EGs affect physical function in older adults.

Since gender distribution was very different among the studies, we performed subgroup analysis based on gender but found no difference in physical or cognitive function between groups at different distributions.

Only 1 study compared EGs versus CEs for depression, and only 1 paper compared the effects of EGs and CEs on QOL. The number of such studies is too small to conduct a meta-analysis. Therefore, future research needs to focus more on older adults’ mental health and further explore and compare the effects of EGs versus CEs on depression and QOL.

According to the literature review, in addition to physical function and mental health, the included studies also compared adherence, motivation, cost-effectiveness, fall rates, accessibility, enjoyment, and attractiveness. One study showed no significant difference in adherence between EGs and CEs [38]; one study showed that the adherence of EGs was significantly higher than that of CEs [56]; and another study showed that EGs can improve the motivation and adherence of older adults in the long-term rehabilitation process [57]. One study found that EGs increased participants’ motivation to do more repetitive movements with minimal or no instruction [54]; another study showed that the presence of motivating stimuli and the novelty aspect of EGs can be particularly important in patients with Parkinson disease who have reduced motivation [57]. One study reported that the extra cost of EGs is minimal compared to CEs (the costs of the computer, screen, safety harness, and platform are relatively low for medium-income countries) [38]. Three studies showed that the fall rate of older adults in the EG group was significantly lower than that in the CE group after the training [38,50,54]. Two studies reported that the advantages of accessibility, enjoyment, and attractiveness of EGs for older adults can further enhance the training effect of CEs [29,49]. However, the number of studies was small enough that a meta-analysis could not be done. We recommend that future research compares these outcomes between EGs and CEs because this information would be invaluable to establishing the added value of EGs versus CEs.

It is worth noting that while EGs may replace CEs in improving physical and cognitive function, it should be considered that not everyone is interested in EGs and not all EGs are suitable for older adults. A recent systematic review and qualitative meta-synthesis conducted by our group explored older adults’
experiences of implementing exergaming programs [70]. We found that a small number of older adults were not interested in EGs, which may be due to age- or health-related factors (eg, vision, hearing, motor skills, or cognitive impairment). At the same time, most older adults have no experience with EGs and worry about whether they can understand and play such games correctly. In addition, people in East Asian countries (such as China, Japan, and South Korea) are more likely to feel embarrassed when using EGs because they feel uncomfortable being observed or judged by others. Finally, most existing EGs are not fully suitable for older adults due to a lack of flexibility and adaptability. However, most of these obstacles can be overcome, for example, by designing older people–friendly EGs for different target groups and giving older adults enough time to train and familiarize themselves with EGs.

This paper has limitations. Since we only included English-language studies, information deviations may occur. As exergaming is a fairly new field, there are few papers on this topic, and the number of studies included (and the total sample size) was not very large; thus, the results should be interpreted with caution. The overall methodological quality of the included studies ranged from medium to excellent, so our findings need to be interpreted with caution. Due to the heterogeneity of intervention types, intervention settings, intervention objects, and measurement tools for some outcome indicators, we were unable to conduct a comprehensive meta-analysis. The existence of publication bias resulted in heterogeneity (eg, differences in the intervention protocol) among the included studies, which reduced the quality of evidence. Finally, it must be acknowledged that the conclusions of this systematic review and meta-analysis may have been influenced by the professional backgrounds of the authors.

Conclusions

Our findings suggest that EGs are a novel and effective strategy for improving physical and cognitive function in older adults. There is no significant difference between EGs and CEs in improving the physical function and cognitive function of older adults, and EGs may replace CEs in these aspects. Our results confirm the effectiveness of EGs in rehabilitation programs for older adults and indicate that EGs may be a novel and feasible alternative to CEs. Future studies should compare the effects of EGs and CEs on cognitive function according to classification of the cognitive status of older adults. Subsequent studies should also quantify the “dose-effect” relationship between EGs and health benefits in older adults, derive optimal intervention doses for EGs, and explore the effects of different types and devices of EGs on physical function in older adults. More high-quality studies with more accurate outcome indicators are needed to further explore and compare the health benefits of EGs versus CEs.

Acknowledgments

We sincerely thank the library of Central South University for its help in the retrieval process of this study. This work was supported by the Special Funding for the Construction of Innovative Provinces in Hunan (grant 2020SK2055), the National Key R&D Program of China (grants 2020YFC2008503 and 2020YFC2008602), the National Natural Science Foundation of China (grant 72004237), the National Key R&D Program of China (2020YFC2008500), and the Changsha County Science and Technology Plan Project (2022037-2).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature retrieval strategy for the PubMed database.

[DOCX File, 14 KB - games_v11i1e42374_app1.docx ]

Multimedia Appendix 2

Characteristics of studies included in this review.

[DOCX File, 20 KB - games_v11i1e42374_app2.docx ]

References


https://games.jmir.org/2023/1/e42374 JMIR Serious Games 2023 | vol. 11 | e42374 | p.621 (page number not for citation purposes)


Abbreviations

CE: conventional exercise
EG: exergame
MMSE: Mini-Mental Status Exam
MoCA: Montreal Cognitive Assessment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
RCT: randomized controlled trial
SF-36: 36 health survey questionnaire
SMD: standardized mean difference
Virtual Reality Technology in Nursing Professional Skills Training: Bibliometric Analysis

Chengang Hong, BSc; Liping Wang, PhD
School of Nursing, Hangzhou Normal University, Hangzhou, China

Corresponding Author:
Liping Wang, PhD
School of Nursing
Hangzhou Normal University
2318 Yuhang Tang Road, Yuhang District
Hangzhou, 311100
China
Phone: 86 15356641080
Email: wangliping111@126.com

Abstract

Background: Nursing professional skills training has undergone significant transformation due to the exponential growth of computer and medical technology. The innovative use of virtual reality (VR) in nursing education has emerged as a cutting-edge technical support technique that has gained attention as a highly effective method for improving nurse training quality.

Objective: This study aims to review the current status of VR technology in nursing professional skills training, research hotspots, and emerging trends in the last 15 years.

Methods: The Web of Science Core Collection database was used to search for literature on VR technology in nursing professional skills training covering the period from 2006 to 2022. Biblioshiny (K-Synth Srl) was used to import and convert the records to Bibliometrix (K-Synth Srl) for analysis, and R (R Core Team) was used for descriptive bibliometric analysis. VOSviewer (Leiden University) was used to cluster co-occurring keywords, and Scimago Graphica (version 1.0.16; Scimago Lab) was used to generate a geographical visualization of published countries and regions.

Results: A total of 1073 papers were analyzed, indicating a surge in research on the application of VR in nursing professional skills training in recent years, as evidenced by a positive trend in annual publication of relevant literature. The majority of studies were from the United States (n=340) and Canada (n=107), and Margaret Verkuyl was the most prolific author, leading the way with 9 publications. Furthermore, “Computerized Virtual Patients in Health Professions Education: a Systematic Review and Meta-Analysis” was the most frequently cited reference. Keywords such as education, simulation, skills, students, and care were most commonly used by researchers.

Conclusions: The bibliometric analysis provides a comprehensive overview of the use of VR in nursing professional skills training, indicating that VR-based training is an effective means of improving the skills and competencies of nursing students and professionals alike. The COVID-19 pandemic has reinforced the importance of developing VR-based distance education, despite challenges such as integrating virtual and real-world training and mitigating safety risks.

Introduction

Background
Nursing work holds a crucial position in health care. Nursing students, serving as a valuable pool of talent for the clinical nursing team, are essential figures in the advancement of nursing practice. They play a pivotal role in promoting the growth and development of the nursing profession and are a crucial component of the health care workforce [1]. Nursing educators face a significant challenge in cultivating the necessary expertise and skills among nursing students to provide effective patient care [2]. Considering the low efficiency of traditional nursing professional skills training methods, it can be challenging to
motivate nursing students to engage with enthusiasm [3]. To ensure that nursing students possess sufficient expertise and are qualified to work in hospitals, many nursing educators use a variety of teaching strategies, including using VR, to enhance the quality of nursing professional skills training.

A virtual reality (VR) system can be defined as a highly interactive 3D digital media environment that generates a simulated reality and creates an immersive experience through the user’s senses of vision, hearing, and touch. Essentially, it provides users with the feeling of being present in a separate virtual environment. The characteristics of VR systems can be summarized into 3 pivotal elements: immersion, interactivity, and sensory perception [4]. VR technology has been extensively applied in nursing theory teaching, nursing skills training, and clinical nursing teaching, both domestically and internationally. Yang and Huang [5] found that traditional nursing skills teaching fails to replicate the actual clinical environment. Students often lack the opportunity to gain genuine sensory experience of nursing operations and may learn various nursing procedures in a conventional and generic manner. This may hinder their ability to effectively transition into clinical work in the early stages of their career. Chu et al [6] showed that a VR-based immersive experiential learning model could be implemented for teaching and that situational learning was achieved through virtual laboratory and simulation training. Gao and Yao [7] provided a new idea for a method to teach endoscopic retrograde cholangiopancreatography to nurses by combining a VR endoscopic simulation system with microteaching. Jung and Park [8] built a head-mounted VR platform to enable students to observe the process of intravenous catheter implantation by modeling the angiography room. These examples provide a compelling argument for the potential of VR technology in enhancing the professional skills of nurses and nursing students.

**Research Problem and Aim**

The concept of “bibliometrics” was proposed by Pritchard in 1969, and it is defined as “the application of mathematical and statistical methods to books and other knowledge dissemination media” [9]. Bibliometrics offers a highly efficient tool for analyzing the progress of scientific research. It enables quantification of information on a specific research topic derived from online scientific citation databases, such as details on authors working in a field, publication numbers, and the distribution of research institutions [10]. Bibliometrics is a powerful tool that can assist in identifying significant literature in a research field and provide valuable data, such as keywords, institutions, country linkages, and distribution characteristics, in the form of a knowledge map. By quantifying the current state and future trends of the research topic, bibliometrics serves as an essential aid in comprehending the development of a particular field [11]. Generally, the greater the number of references included in a bibliometric method, the better we can comprehend the landscape of a research field [12]. A scoping review conducted by Yang and Huang [5] outlines the applications of VR technology in nursing professional skills training. There has been limited research on the scientific output of this field using bibliometric and visualization analyses. As a result, it is critical to comprehend the research status and hot spots of this field to aid in the advancement of nursing professional skills for nurses and nursing students. This study strives to offer a comprehensive analysis of the research field using the Web of Science (WOS) Core Collection as a basis and to serve as a valuable reference for future research.

**Methods**

**Data Sources and Search Strategy**

For this study, articles and reviews that were published in English between 2001 and 2021 were included, using the WOS Core Collection database as the primary data source: (1) the retrieval expression was constructed using a formula that used a Boolean logic operator: TS=(virtual reality OR VR OR virtual medicine OR augment reality OR mixed reality OR virtual simulation) AND TS =(nursing* OR nursing practice OR nursing training OR academic training OR serious game), and the scope of literature was restricted to only articles and reviews; (2) the English language was selected; (3) there were 1073 documents retrieved in the search of WOS (as updated on 11 November 2022); and (4) all records were exported to plain text files, including the full record and cited references.

**Bibliometric Analysis**

A description of the bibliometric analysis method can be found in Aria and Cuccurullo [13]. Five rigorous steps, namely study design, data collection, data analysis, data visualization, and interpretation, were included in the analysis [14]. The complete methodology is schematically illustrated in Figure 1. The initial stage was the study design phase; nursing professional skills training was selected as the study topic. The research data resource chosen was the WOS Core Collection database, from which literature retrieval returned 1242 documents during the data collection phase. The review followed the structured approach outlined by Kable et al [15] for researching and evaluating literature. In this study, 2 reviewers (CH and LW) conducted an independent screening of the search results, with cross-checking used to resolve any disagreements. Training in document retrieval and screening was received by all researchers based on the textbook Medical Literature Information Retrieval [13]. A document-type filter was applied to the WOS, and articles and data papers were included, resulting in a sample of 1073 papers published between 2006 and 2022. All records were imported to Biblioshiny (K-Synth Srl) and converted to Bibliometrix (K-Synth Srl) for subsequent analysis. In the third stage, R (R Core Team) was used for descriptive bibliometric analysis and to develop a matrix of all documents. Biblioshiny, the Tidyverse package ggplot2 (Hadley Wickham), and VOSviewer (Leiden University) were used in the fourth stage to create a keyword cluster map. Scimago Graphica (version 1.0.16; Scimago Lab) was used to generate geographic visualization maps of the countries or regions. To identify influential sources, the Bradford law was used to unveil the journal distribution. The Methods section presents an interpretation of the data analysis and visualization outcomes.
**Ethical Considerations**

An ethical approval application was not filed for this paper. The Chinese Hospital Association has stated that an ethics review is unnecessary for secondary analyses of published data.

**Results**

**Descriptive Bibliometric Analysis**

Each period’s publication count is a reflection of the research field’s development trends. Figure 2 presents the scientific output over the course of the study. There were 3 periods based on the number of publications: 2006 to 2013, 2014 to 2019, and 2019 to 2022. As seen in Figure 2, no relevant papers were released before 2006. From 2006 to 2013, the number of published articles increased annually, but at a relatively sluggish pace. The quantity of publications in 2019 (n=76) surpassed 75 for the first time. The number of articles published between 2019 and 2021 increased significantly, indicating that more scholars were focusing on the potential of VR in the nursing professional skills training field. Table 1 provides essential details on the 1073 research papers published in the WOS Core Collection database from 2006 to 2022. The annual average number of published research papers was 61.29, with each paper having an average citation count of 11.52. These papers had 4369 authors, with 59 of them being single-author papers. The average number of authors for each research paper was 2 (2.33). The collaboration index, which is the total number of authors of multi-authored papers divided by the total number of multi-authored papers, was 2.38. These papers also produced a total of 2388 author keywords.
The production of research on virtual reality–based nursing professional skills training in the scientific domain between 2006 and 2022.

![Graph showing the production of research on virtual reality–based nursing professional skills training](image)

Table 1. Crucial information on virtual reality–based nursing professional skills training literature determined by the bibliometric analysis.

<table>
<thead>
<tr>
<th>Information</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents, n</td>
<td>1073</td>
</tr>
<tr>
<td>Sources(^a) (frequency distribution), n</td>
<td>300</td>
</tr>
<tr>
<td>Timespan (years)</td>
<td>2006-2022</td>
</tr>
<tr>
<td>References, n</td>
<td>30,766</td>
</tr>
<tr>
<td>Author keywords, n</td>
<td>2388</td>
</tr>
<tr>
<td>Keywords Plus(^b), n</td>
<td>1469</td>
</tr>
<tr>
<td>Authors, n</td>
<td>4369</td>
</tr>
<tr>
<td>Authors’ appearances (frequency distribution), n</td>
<td>5283</td>
</tr>
<tr>
<td>Authors of single-authored documents, n</td>
<td>59</td>
</tr>
<tr>
<td>Authors of multi-authored documents, n</td>
<td>1014</td>
</tr>
<tr>
<td>Documents per author (average), n</td>
<td>0.246</td>
</tr>
<tr>
<td>Coauthors per document (average), n</td>
<td>4.92</td>
</tr>
<tr>
<td>Citations per document (average), n</td>
<td>11.52</td>
</tr>
<tr>
<td>Collaboration index</td>
<td>0.23</td>
</tr>
</tbody>
</table>

\(^a\)Sources include journals and books.

\(^b\)Total number of phrases that frequently appear in the title of an article’s references.

The WOS Research Areas

One or more subject categories (SCs) are linked to each publication indexed in the WOS. In this study, the number of research areas covered by VR-based nursing professional skills training education literature increased from 2 in 2006 to 22 in 2022 (Figure 3). Figure 4 displays the annual changes in productivity in the top 10 areas of VR-based nursing professional skills training education, which illustrates changes in the focus areas of VR-based nursing professional skills training education. The most popular research area was nursing, followed by education and educational research and health care sciences and services. With the increasing maturity of VR technology and stable production of VR equipment, the explosive growth in publications in the field of nursing professional skills training can be attributed to the use of various teaching strategies, including VR, by many nursing educators to enhance the quality of education.
Figure 3. The count of Web of Science research areas included in the literature of VR-based nursing professional skills training.

Figure 4. Yearly variations in the productivity of the top 10 Web of Science research areas in specific fields.
**Research Countries and Institutions**

Table 2 highlights that the United States has contributed the highest number of publications (340/1042, 32.63%) in the domain of VR-based nursing professional skills training, followed by Canada (107/1042, 10.27%), China (89/1042, 8.54%), Australia (71/1042, 6.81%), and the United Kingdom (51/1042, 4.89%). Of the total number of articles, 64.65% (658/1042) were published in these 5 countries. To generate a geographic visualization map (Figure 5), countries with more than 5 publications were analyzed using Scimago Graphica. The lines on the map indicate the degree of cooperation between different countries, with the thickness of the lines reflecting the extent of cooperation. It is evident from the map that the United States, Canada, Australia, China, and other countries have actively collaborated with other countries. Additionally, several scholars from Europe have shown interest in VR-based nursing professional skills training education and have conducted joint research projects to a certain extent. Table 3 highlights that while the United States has the highest number of publications, Canada’s top 10 institutions rank higher in total citations, suggesting that both countries have robust support and funding for VR-based nursing professional skills training research.

**Table 2.** Countries with the highest total citation count for virtual reality–based nursing professional skills training research, ranked by the top 10.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Country</th>
<th>Articles, n</th>
<th>Citations, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States</td>
<td>340</td>
<td>4215</td>
</tr>
<tr>
<td>2</td>
<td>Canada</td>
<td>107</td>
<td>1186</td>
</tr>
<tr>
<td>3</td>
<td>China</td>
<td>89</td>
<td>867</td>
</tr>
<tr>
<td>4</td>
<td>Australia</td>
<td>71</td>
<td>1145</td>
</tr>
<tr>
<td>5</td>
<td>United Kingdom</td>
<td>51</td>
<td>1027</td>
</tr>
<tr>
<td>6</td>
<td>Brazil</td>
<td>49</td>
<td>227</td>
</tr>
<tr>
<td>7</td>
<td>Korea</td>
<td>42</td>
<td>181</td>
</tr>
<tr>
<td>8</td>
<td>Spain</td>
<td>35</td>
<td>309</td>
</tr>
<tr>
<td>9</td>
<td>Singapore</td>
<td>30</td>
<td>489</td>
</tr>
<tr>
<td>10</td>
<td>Turkey</td>
<td>28</td>
<td>115</td>
</tr>
</tbody>
</table>

**Figure 5.** Map displaying research collaboration between countries.
Table 3. The institutions ranked in the top 10 based on the total number of citations for virtual reality–based nursing professional skills training.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Country</th>
<th>Articles, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Toronto</td>
<td>Canada</td>
<td>42</td>
</tr>
<tr>
<td>National University of Singapore</td>
<td>Singapore</td>
<td>38</td>
</tr>
<tr>
<td>Université de Montréal</td>
<td>Canada</td>
<td>23</td>
</tr>
<tr>
<td>Toronto Metropolitan University</td>
<td>Canada</td>
<td>22</td>
</tr>
<tr>
<td>Centennial College</td>
<td>Canada</td>
<td>20</td>
</tr>
<tr>
<td>Johns Hopkins University</td>
<td>USA</td>
<td>20</td>
</tr>
<tr>
<td>University of Ottawa</td>
<td>Canada</td>
<td>19</td>
</tr>
<tr>
<td>University of São Paulo</td>
<td>Brazil</td>
<td>19</td>
</tr>
<tr>
<td>Nanyang Technological University</td>
<td>Singapore</td>
<td>18</td>
</tr>
<tr>
<td>University of Almeria</td>
<td>Spain</td>
<td>18</td>
</tr>
</tbody>
</table>

Most Influential Source Journals

Table 4 shows that Clinical Simulation in Nursing was the most cited journal with 1401 of 30,766 total citations (4.55%), while the Journal of Medical Internet Research, with an impact factor of 7.077, was the most impactful among the top 10 academic journals. Half of the journals in the table were ranked in the top quartile position. We also analyzed the distribution of research papers among leading sources and found that the top 5 journals accounted for 304 of 1042 (29.17%) of the total number of papers. The top 5 journals with the highest numbers of papers published were Clinical Simulation in Nursing (n=125), Nurse Education Today (n=93), Journal of Nursing Education (n=32), Journal of Medical Internet Research (n=27), and Nurse Education in Practice (n=27), with Clinical Simulation in Nursing having the most local citations (Table 4). According to the Bradford law, the source journals of VR-based nursing professional skills training papers were widely scattered, and the top 10 most influential journals were selected based on the number of local citations. Journals that were identified as core source journals in the field of VR-based nursing professional skills training according to the Bradford law are marked in Figure 6. These included Clinical Simulation in Nursing, Nurse Education Today, Journal of Nursing Education, Journal of Medical Internet Research, Nurse Education in Practice, CIN-Computers Informatics Nursing, and Journal of Advanced Nursing. These journals played a crucial role in the development of VR-based nursing professional skills training during the study period.

Table 4. Ranking of the top 10 journals in VR-based nursing professional skills training based on the number of local citations.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Local citations, n</th>
<th>Documents, n</th>
<th>Impact factor(^{b})</th>
<th>Rank (quartile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Simulation in Nursing (^{a})</td>
<td>1401</td>
<td>125</td>
<td>2.856</td>
<td>Q1</td>
</tr>
<tr>
<td>Nurse Education Today (^{a})</td>
<td>1397</td>
<td>93</td>
<td>3.906</td>
<td>Q1</td>
</tr>
<tr>
<td>Journal of Nursing Education (^{a})</td>
<td>526</td>
<td>32</td>
<td>2.381</td>
<td>Q2</td>
</tr>
<tr>
<td>Journal of Medical Internet Research (^{a})</td>
<td>470</td>
<td>27</td>
<td>7.077</td>
<td>Q1</td>
</tr>
<tr>
<td>Nurse Education in Practice (^{a})</td>
<td>440</td>
<td>27</td>
<td>3.43</td>
<td>Q1</td>
</tr>
<tr>
<td>CIN-Computers Informatics Nursing (^{a})</td>
<td>421</td>
<td>25</td>
<td>2.146</td>
<td>Q3</td>
</tr>
<tr>
<td>Journal of Advanced Nursing (^{a})</td>
<td>407</td>
<td>19</td>
<td>3.057</td>
<td>Q1</td>
</tr>
<tr>
<td>Nurse Educator</td>
<td>385</td>
<td>18</td>
<td>2.791</td>
<td>Q1</td>
</tr>
<tr>
<td>JMIR Serious Games</td>
<td>382</td>
<td>18</td>
<td>3.364</td>
<td>Q2</td>
</tr>
<tr>
<td>Nursing Education Perspectives</td>
<td>343</td>
<td>17</td>
<td>0.59</td>
<td>Q3</td>
</tr>
</tbody>
</table>

\(^{a}\)Journals classified as core resources in virtual reality–based nursing professional skills training according to the Bradford law.

\(^{b}\)Impact factor in 2021.
Figure 6. Temporal analysis of the publication sources of virtual reality–based nursing professional skills training.

### Most Influential Authors

The H index is a widely accepted measure of scientific performance based on the number of citations received by a scientist’s published papers [16]. Table 5 shows the top 10 authors with the largest H index: Margaret Verkuyl (H index 9), Shujuan Liao (H index 7), Daria Romaniuk (H index 7), Hilde Eide (H index 6), Sharon L Farra (H index 6), Michelle Hughes (H index 6), Paula Mastrilli (H index 6), Michelle Aebersold (H index 5), Lynda Atack (H index 5), and Luciana Mara Monti Fonseca (H index 5). The researcher with the highest number of citations, indicating significant influence, was Margaret Verkuyl. She is also the author with the highest density of collaboration (Figure 7). The top 10 most influential researchers included researchers from Canada (n=5), China (n=1), Norway (n=1), Brazil (n=1), and the United States (n=2). The 1073 papers involved 4369 authors. In total, 59 authors published 65 documents as sole authors, with an average of 4.92 co-authors per document and a collaboration index of 0.23. Each author contributed an average of 0.246 documents to the field, and on average, each document had 1.1 authors and 4.92 coauthors. These findings suggest that VR-based nursing professional skills training research is often a collaborative effort involving multiple authors.
Table 5. Authors ranked by the H index, with the top 10 being the most influential.

<table>
<thead>
<tr>
<th>Author</th>
<th>H Index</th>
<th>G Index</th>
<th>Times cited(^a), n</th>
<th>Scientific productions, n</th>
<th>First year published</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret Verkuyl</td>
<td>9</td>
<td>16</td>
<td>268</td>
<td>21</td>
<td>2016</td>
<td>Canada</td>
</tr>
<tr>
<td>Shujuan Liao</td>
<td>7</td>
<td>13</td>
<td>250</td>
<td>13</td>
<td>2014</td>
<td>China</td>
</tr>
<tr>
<td>Daria Romaniuk</td>
<td>7</td>
<td>9</td>
<td>183</td>
<td>9</td>
<td>2016</td>
<td>Canada</td>
</tr>
<tr>
<td>Hilde Eide</td>
<td>6</td>
<td>10</td>
<td>124</td>
<td>10</td>
<td>2011</td>
<td>Norway</td>
</tr>
<tr>
<td>Sharon L. Farra</td>
<td>6</td>
<td>7</td>
<td>114</td>
<td>7</td>
<td>2013</td>
<td>United States</td>
</tr>
<tr>
<td>Michelle Hughes</td>
<td>6</td>
<td>9</td>
<td>147</td>
<td>9</td>
<td>2017</td>
<td>Canada</td>
</tr>
<tr>
<td>Paula Mastrilli</td>
<td>6</td>
<td>10</td>
<td>165</td>
<td>10</td>
<td>2016</td>
<td>Canada</td>
</tr>
<tr>
<td>Michelle Aebersold</td>
<td>5</td>
<td>5</td>
<td>126</td>
<td>5</td>
<td>2015</td>
<td>United States</td>
</tr>
<tr>
<td>Lynda Attack</td>
<td>5</td>
<td>7</td>
<td>137</td>
<td>7</td>
<td>2016</td>
<td>Canada</td>
</tr>
<tr>
<td>Luciana Mara Monti Fonseca</td>
<td>5</td>
<td>8</td>
<td>68</td>
<td>8</td>
<td>2014</td>
<td>Brazil</td>
</tr>
</tbody>
</table>

\(^a\)Web of Science Core Collection times cited count.

**Figure 7.** Visualization map of author collaboration in research on virtual reality–based nursing professional skill training.

**Most Influential Papers**

This section provides a list of the most influential papers (based on the number of citations) published between 2006 and 2022. Cocitations within a group of authors suggest that certain documents may include concept symbols. It is noteworthy that Tables 6 and 7 show a difference between the local citation score (LCS), which reflects the number of citations within the field, and the global citation score (GCS), which represents the total number of citations in WOS. The paper with greatest influence according to LCS was “Clinical Virtual Simulation in Nursing Education: Randomized Controlled Trial” (n=49). The paper with greatest influence according to GCS was “Computerized Virtual Patients in Health Professions Education: a Systematic Review and Meta-Analysis” (n=273). The papers in question may encompass theories that form the foundation of the field, early works that lay groundwork, and methodological principles [17].
Table 6. Top 10 papers ranked by the local citation score.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Journal</th>
<th>Local citation score</th>
<th>Global citation score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Padilha et al [18]</td>
<td>2019</td>
<td>J Med Internet Res</td>
<td>49</td>
<td>108</td>
</tr>
<tr>
<td>Foronda et al [19]</td>
<td>2020</td>
<td>Simul Healthc</td>
<td>49</td>
<td>87</td>
</tr>
<tr>
<td>Cobbett and Snelgrove-Clarke [20]</td>
<td>2016</td>
<td>Nurs Educ Today</td>
<td>43</td>
<td>62</td>
</tr>
<tr>
<td>Butt et al [22]</td>
<td>2018</td>
<td>Clin Simul Nurs</td>
<td>35</td>
<td>74</td>
</tr>
<tr>
<td>Cook et al [23]</td>
<td>2010</td>
<td>Acad Med</td>
<td>33</td>
<td>273</td>
</tr>
<tr>
<td>Verkuyl et al [26]</td>
<td>2017</td>
<td>Clin Simul Nurs</td>
<td>26</td>
<td>50</td>
</tr>
<tr>
<td>Kidd et al [27]</td>
<td>2012</td>
<td>J Psychosoc Nurs Men</td>
<td>26</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 7. Top 10 papers ranked by the global citation score.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Journal</th>
<th>Local citation score</th>
<th>Global citation score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook et al [23]</td>
<td>2010</td>
<td>Acad Med</td>
<td>33</td>
<td>273</td>
</tr>
<tr>
<td>Cook et al [28]</td>
<td>2013</td>
<td>Acad Med</td>
<td>3</td>
<td>193</td>
</tr>
<tr>
<td>Zendejas et al [29]</td>
<td>2013</td>
<td>J Gen Intern Med</td>
<td>1</td>
<td>173</td>
</tr>
<tr>
<td>Gentry et al [31]</td>
<td>2019</td>
<td>J Med Internet Res</td>
<td>22</td>
<td>135</td>
</tr>
<tr>
<td>Kipping et al [32]</td>
<td>2012</td>
<td>Burns</td>
<td>6</td>
<td>130</td>
</tr>
<tr>
<td>Triola et al [33]</td>
<td>2006</td>
<td>J Gen Intern Med</td>
<td>13</td>
<td>123</td>
</tr>
<tr>
<td>Padilha et al [18]</td>
<td>2019</td>
<td>J Med Internet Res</td>
<td>49</td>
<td>108</td>
</tr>
<tr>
<td>Robert et al [34]</td>
<td>2014</td>
<td>Front Aging Neurosci</td>
<td>1</td>
<td>107</td>
</tr>
<tr>
<td>Manera et al [35]</td>
<td>2015</td>
<td>Front Aging Neurosci</td>
<td>3</td>
<td>106</td>
</tr>
</tbody>
</table>

Distribution of Keywords

By conducting data cleaning to eliminate coding errors that could arise from different articles using varying keywords to introduce the same concept, we identified 2388 author keywords in 1073 published papers on VR-based nursing professional skills training between 2006 and 2022. The keyword analysis enables summary of the study’s topics in the field and exploration of hotspots. For instance, nursing and nurses were recorded as Nurses to prevent discrepancies. Education had the highest frequency of 178 among the keywords, with other high-frequency keywords being Simulation (n=148), Skills (n=95), Students (n=84), Care (n=77), Impact (n=64), Knowledge (n=64), Nurses (n=62), Virtual-reality (n=58), and Technology (n=47), as shown in Table 8. Additionally, Multimedia Appendix 1 shows a keyword cluster map created using VOSview, while Figure 8 depicts the thematic evolution of the keywords.
Table 8. Top 10 keywords related to the application of virtual reality in nursing professional skills training.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Keyword</th>
<th>Frequency</th>
<th>Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Education</td>
<td>178</td>
<td>0.06</td>
</tr>
<tr>
<td>2</td>
<td>Simulation</td>
<td>148</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>Skills</td>
<td>95</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>Students</td>
<td>84</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>Care</td>
<td>77</td>
<td>0.05</td>
</tr>
<tr>
<td>6</td>
<td>Impact</td>
<td>64</td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>Knowledge</td>
<td>64</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>Nurses</td>
<td>62</td>
<td>0.02</td>
</tr>
<tr>
<td>9</td>
<td>Virtual-reality</td>
<td>58</td>
<td>0.03</td>
</tr>
<tr>
<td>10</td>
<td>Technology</td>
<td>47</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Figure 8. Map showing thematic evolution of keywords.

Discussion

General Information

Based on the annual number of publications, the publication years can be divided into 3 phases: 2006 to 2013, 2014 to 2019, and 2019 to 2022. The first phase, spanning from 2006 to 2013, saw an increase in the number of articles, albeit at a relatively slow growth rate. The second phase, from 2014 to 2019, was marked by significant advancements in VR device and software technology. This was exemplified by the launches of Project Glass by Google in 2012 and the first generation of the Oculus Rift VR headset by Oculus. These improvements resulted in an increase in the availability and quality of VR components [17]. In the period between 2014 and 2021, there was a noticeable increase in interest among scholars in computer science, as evidenced by periods 2 and 3. In the most recent phase, from 2019 to 2021, the number of publications exceeded 75 for the first time, with 76 articles published. This trend indicates an expanding interest among scholars in exploring the potential of VR technology in the field of nursing professional skills training.

Contributions to VR-based nursing professional skills training came from various countries worldwide, including the United States, Canada, China, Australia, the United Kingdom, Brazil, Korea, Spain, Singapore, and Turkey, with some taking a more prominent role. Notably, China is one of the few Asian countries to contribute to this field, indicating a lack of research emphasis on this topic among Asian scholars. In contrast, Canadian institutions and scholars have demonstrated continued interest in research in this area. A bibliometric analysis of VR and nursing professional skills training revealed Canada to have the highest contribution among all countries [23].
The frequency of cocitation can serve as a reflection of a journal’s quality and influence. Among the ten journals with the highest cocitation frequency, half were found to be ranked in the top quartile. This suggests significant impact and recognition in the international community, indicating that the application of VR in nursing professional skills training has gained worldwide attention.

Following the analysis of cocited papers, it was found that “Clinical Virtual Simulation in Nursing Education: Randomized Controlled Trial” (n=49) was the most influential paper in terms of LCS. This study may contain foundational theories, pioneering early work, and methodological principles in the field. Similarly, “Computerized Virtual Patients in Health Professions Education: a Systematic Review and Meta-Analysis” (n=273) was deemed the most influential paper in terms of GCS. These works demonstrate how innovative research can impact an academic discipline for years to come and continue to inform contemporary research in the field.

Research Hot Spots and Frontiers

By using a combination of high-frequency keyword statistics, visualization analysis, and including relevant literature, this study indicates that an overwhelming majority of both domestic and international researchers have concentrated their efforts on exploring the application of VR technology in nursing education with a particular emphasis on nursing professional skills training.

Owing to variations in technological advancement, each country and region has conducted research to its own specifications with regards to the use of VR in nursing education. As an illustration, the VR nursing skills training system can provide a significant contribution to ensuring the proficiency of nursing students by enabling them to learn essential operational skills without the need for physical attendance, thereby providing them with the opportunity to review training materials anywhere at any time [23]. International research in this field commonly uses desktop VR system software to enhance students’ clinical thinking skills. This type of software emphasizes immersive patient scenarios presented on the computer screen in which students are “placed” in a simulated clinical setting to observe and address the patient’s discomfort and pain. The aim is to develop students’ proficiency in making informed clinical decisions based on complex patient conditions [36]. The VR system can be improved in a short time according to teaching and learning needs, quickly meeting the skills training and teaching needs of various courses and improving the training effect. In order to give full play to the maximum benefit of the virtual system, it is suggested that schools and hospitals with suitable conditions choose VR technology to carry out skills training for students and promote their early adaptation to clinical practice.

In the field of nursing education, there are notable technical issues that exist with the application of VR technology at present. Caylor et al [37] pointed out that VR systems are faced with the problem of unclear navigation mechanisms within the operation interface. Poor experiences among nursing students during simulation learning are attributed to the confusing nature of the navigation interface within VR software platforms. Wilson et al [38] suggested that in future research on VR technology more attention be given to creating a simulation system with high fidelity, increasing the complexity of the cases presented within VR and reproducing challenging clinical case scenarios to enhance subjects’ diagnostic reasoning abilities. Moreover, the development level and quality of VR software must be improved. This improvement is necessary due to the lack of understanding among VR technicians of the medical care field, which leads to failures to accurately reflect the key points and difficulties that teachers aim to emphasize. Additionally, it is vital to combine the advantages of VR teaching with existing nursing practices to ensure that VR technology is used effectively in nursing education [39]. To maximize the potential of VR technology in nursing education, researchers must consider providing a more comprehensive VR technology intervention, emphasizing multidisciplinary cooperation, and enhancing the familiarity of software technology developers with the nursing field. Additionally, teachers’ proficiency in the operation of the virtual teaching process must be improved. To better explore the efficacy of VR technology in nursing education, it is crucial to enhance the VR teaching environment to improve nursing students’ ability to apply the skills learned in virtual settings to clinical practice.

Applications and Development Trends in Research

Over the last 15 years, VR technology has been extensively implemented in nursing professional skills training. As mobile computing and sophisticated software programs continue to develop, highly immersive VR systems have become increasingly affordable and accessible [40]. Popular VR systems currently used in nursing professional skills training include Kinect from the United States, Nintendo Wii from Japan, HTC Vive from Taiwan, China, and Samsung Gear VR from South Korea. By integrating these VR systems into nursing education, we aim to facilitate the development of students’ initiative and promote self-directed learning abilities, ultimately enhancing their clinical practice acumen.

As COVID-19 continues to impact the world, students’ access to skills training may be limited. However, VR technology offers a viable solution to this problem. VR-based nursing skills training immerses students in a simulated environment, offering an opportunity to develop critical skills despite the pandemic. According to research, the use of a VR immersive experiential learning model for teaching fosters situational teaching through virtual laboratory and simulation training [6]. Furthermore, Gao and Yao [7] combined use of a VR endoscopic simulation system with the microteaching method to provide a new idea for teaching nurses endoscopic retrograde cholangiopancreatography. Jung and Park [8] developed a head-mounted VR platform for observing the process of intravenous catheter implantation in a highly realistic sensory scenario. This platform enabled students to view the entire arterial and venous system of the body, strengthening their theoretical and practical operation skills. Zhang et al [41] developed a highly realistic scenario simulation system for improving teachers’ teaching ability and cultivating students’ initiative, enthusiasm, and autonomous learning ability.

To establish VR technology as a reliable tool for improving nursing professional skills training, researchers must integrate knowledge from various fields, including nursing, computer
science, neuroscience, psychology, social cognition, multisensory perception, and multimedia development. This requires an interdisciplinary approach that takes into account the needs of all stakeholders, including students, teachers, nurses, and software developers [42]. Thus, research on VR-based nursing professional skills training must involve a broad range of individuals to create effective teaching strategies that meet the unique needs of every student. High-quality research is necessary to explore the full potential of VR technology across various disciplines.

**Strengths and Limitations**

This is the first bibliometric analysis of VR-based professional skills training in nursing, providing a comprehensive understanding of the current state of VR in nursing education. Through analyzing keywords, countries or regions, institutions, authors, journals, and references, we have gained valuable insight into the research landscape. In the future, longitudinal in-depth studies on current hot topics or large, multicenter, high-quality randomized controlled studies guided by research trends can further enhance this field. At present, research in this field is predominantly limited due to technology and cost. To promote growth, policy and financial support should be provided to encourage researchers to learn from foreign advanced experience and enhance VR technology. Nursing staff must participate in program design from a nursing education perspective to create more suitable VR equipment and procedures. By vigorously developing the application of VR in the field of nursing education, we can better prepare our students for clinical practice.

However, it is important to note that this study used only one database (WOS), which may have excluded high-quality papers in the field. Additionally, the current word segmentation algorithm used in bibliometrics lacks necessary intelligence, leading to inaccurate extraction of some keywords. Thus, future bibliometric studies must strive to enhance the semantic understanding of citation data to facilitate accurate word segmentation and intelligent extraction of bibliometric knowledge. To ensure a more comprehensive and precise theoretical basis, future research should expand the database and source high-quality papers both domestically and internationally. This will better promote the advancement of nursing education and practice to new levels.

**Conclusions**

To review the application of VR technology in nursing professional skills training, this study used the WOS database, R, Scimago Graphica, and VOSview to provide a scientific and intuitive analysis. Bibliometric analysis indicates that research on VR-based professional skills training in nursing has experienced rapid development over the past 15 years. This is evident from the rising number of publications in core journals and increased collaboration between authors and countries or regions. Although this paper demonstrates an exponential growth trend in VR-based nursing professional skills training and positive feedback in the field's advancement, further investigation is required to understand the feedback mechanism.

**Acknowledgments**

Funding support for this study was provided by the Natural Science Foundation of Zhejiang Province, China (LY18H160061; 2017).

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Map showing research collaboration between keywords. [PNG File, 857 KB - games_v11i1e44766_app1.png]

**References**


Abbreviations

- GCS: global citation score
- LCS: local citation score
- SC: subject category
- VR: virtual reality
- WOS: Web of Science
Dimensions of Interactive Pervasive Game Design: Systematic Review

Liu Kai1,2*, PhD; Wee Hoe Tan3,4*, PhD; Erni Marlina Saari5*, PhD

1School of Art and Design, Henan University of Engineering, Zhengzhou, China
2Faculty of Art, Sustainability and Creative Industry, Sultan Idris Education University, Tanjong Malim, Malaysia
3De Institute of Creative Arts and Design, UCSI University, Kuala Lumpur, Malaysia
4Serious Games Association, Singapore, Singapore
5Faculty of Computing and Meta-Technology, Sultan Idris Education University, Tanjong Malim, Malaysia

*all authors contributed equally

Corresponding Author:
Liu Kai, PhD
School of Art and Design
Henan University of Engineering
No. 1, Xianghe Road, Xinzheng
Zhengzhou, 451191
China
Phone: 86 13298339393
Fax: 86 62503826
Email: liukai@haue.edu.cn

Abstract

Background: As the gaming industry grows around the world, playing pervasive games is becoming an important mode of entertainment. A pervasive game is one in which the game experience extends into the actual world or where the fictive world of the game merges with the physical world. How pervasive games can adapt to the ever-changing nature of technology and design in current society requires a comprehensive review.

Objective: In this systematic review, we aimed to measure and analyze 4 dimensions of pervasive games through development, technology, experience, and evaluation. Moreover, we also aimed to discover and interpret their relationship with game, interaction, experience, and service design.

Methods: We first chose 3 well-known databases, Web of Science, Scopus, and EBSCO, and searched from 2013 to April 2022. A strictly thorough Boolean search for research keywords such as “pervasive game,” “design,” and “interactive” resulted in 394 relevant articles. These articles were identified, screened, and checked for eligibility to find valid and useful articles, which were then categorized and analyzed using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method.

Results: The systematic selection was finally left with 40 valid and valuable articles. After categorization and analysis, all articles were classified according to 4 main themes, which were design and development (11/40, 28%), interaction and technology (15/40, 38%), users and experience (9/40, 23%), and evaluation and service (5/40, 13%). These 4 main areas can be subdivided into several smaller areas.

Conclusions: In the 4 areas of game design, interaction design, experience design, and service design, many scholars have studied pervasive games and made contributions. Although the development and technology of pervasive games have evolved with the times, there is still a need to strengthen emerging design concepts within pervasive games.

(JMIR Serious Games 2023;11:e42878) doi:10.2196/42878

KEYWORDS
interactive; pervasive game; systematic review; design; mobile phone
**Introduction**

**Background**

During the global COVID-19 pandemic, people from all walks of life played video games to find joy, connection, and a sense of belonging. A game is a source of entertainment and comfort for millions of people in all countries, and it transcends age, race, gender, platform, and political party. In 2021, according to the Entertainment Software Association survey, 90% of gamers agreed that video games bring joy and strongly agreed that they can inspire (79%), provide mental stimulation (87%), and relieve stress (87%) [1]. Sensor Tower released its *Mobile Gaming* 2022 report, which provides insight into how challenges affect different regions and game genres and which markets continue to grow in the face of these challenges. The biggest game opportunity is in China, as can be interpreted from the data. China currently leads the mobile game industry in terms of revenue and downloads and is expected to maintain its dominant position in the coming years [2]. Multiplayer online battle arena has proven to be one of the most challenging but rewarding game genres in the mobile game market, bringing a series of huge successes and painful failures. The latest high-profile multiplayer online battle arena to be released on cell phones is *Pokémon Unite*, developed by *Pokémon Company* in partnership with Tencent’s TiMi Studio. *Pokémon GO* was downloaded more than 15 million times in the first 2 days of its release, making it the biggest release ever. By the end of the first week, the game had reached approximately 30 million downloads in 7 days [3].

Pervasive games are a very important subfield among the many game classifications. As early as 2009, Montola et al [4] wrote in their book that pervasive games were defined as having *one* or *more* salient features that expand the contractual magic circle of play spatially, temporally, or socially. Initially, pervasive games were relatively simple to play. Game players with other players and nonplayer characters communicate via phone, email, or MSN Messenger [4]. However, now, pervasive games are a whole new game style that extends game experiences into the actual world, weaving into the fabric of players’ real-life settings [5]. According to some researchers, a pervasive game is when the player’s experience is extended to the actual world, achievable because of the device’s sensors [6]. These games are a relatively new type of entertainment that takes the game experience out of the device and into the actual environment, merging virtual and physical realities [7].

Pervasive games use mobile devices’ data collection, geotagging, e-commerce, internet cookies, and social media capabilities [8]. Pervasive games rose to popularity with the release of Niantic’s *Pokémon GO* in 2016. Two games now dominate the pervasive games market: *Pokémon GO* and *Ingress* [9]. Ingress has a rather simple core loop and low polish, but its dynamics seem to be tailored to enforce team socializing and collaboration [10]. After the successor to *Ingress, Pokémon GO* offers a more complicated primary gameplay loop. The mechanics of *Pokémon GO*’s main gameplay loop are more intertwined than those of *Ingress*; nonetheless, the game focuses significantly more on the individual’s experiences in the Pokémon world and far less on social connections and collaboration [11].

**Objectives**

On the basis of the abovementioned background retrospect, it can be seen that further research on pervasive games is necessary. As design continues to evolve, many new concepts are being added. There is a need to add more current design factors in the game design industry. For example, interaction design, experience design, service design, and branding design are popular design concepts that constantly influence game design. In addition, the existing systematic review articles fail to provide detailed information on the review process. This includes keyword identification, article screening, and article eligibility. Moreover, because of this situation, prospective researchers could not reconstruct the survey, authorize interpretation, or assess the breadth of the data. Furthermore, this study is important because it provides researchers with an understanding of peer literature reviews and helps researchers better understand pervasive game design issues that may require academic attention. The current systematic analysis was conducted to answer the primary research question: Does the pervasive game embody other conceptions of thinking in the design field? The survey’s primary focus was on perceptions of pervasive game design, including perceptions of design thinking in interaction design, experience design, service design, and branding design.

The *Methods* section describes the procedure used to answer the research topic of this study. The *Results* section conducts a systematic review and synthesis of the scientific literature to identify, select, and analyze the research needed on pervasive game design within design thinking situations. Finally, the *Discussion* section discusses the steps that must be followed, emphasizing the importance of aspiring academics understanding the issues.

**Methods**

**Overview**

Several studies related to system evaluation have been conducted globally. However, only a few studies have focused on the design thinking connotations of pervasive games in the context of design thinking [4-8]. This section describes the need for a systematic analysis of the design thinking connotations of pervasive games. In contrast, the following section presents the methodology used to find the research answers formulated by this study. This literature analysis is divided into four sections: (1) design and development, (2) interaction and technology, (3) user and experience, and (4) evaluation and service. Furthermore, this section systematically reviews and synthesizes the scientific literature to distinguish, screen, and analyze important pervasive game research. Finally, reflecting on potential scholars suggests what actions should be taken to address the issues raised. The prerecording systematic reviews and meta-analysis (PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses]) approach was used in this analysis, a documented standard for conducting a systematic literature review. In general, publication rules are required to aid writers in assessing and reviewing the quality

---

https://games.jmir.org/2023/1/e42878
and rigor of a review with relevant and necessary details. PRISMA also emphasizes the randomized study evaluation survey, an important feature in systematic analysis reports for various study types [12] (Figure 1).

In terms of instruments, the resilient character of 3 electronic databases, Scopus, Web of Science, and EBSCO, was used to evaluate this research methodology. This review’s electronic database search included articles relevant to education, information technology, and social science. However, similar to Scopus, no database is perfect and detailed [13]. They mainly cover various fields, such as business and industry, economics, information technology, humanities, social sciences, communication and dissemination, education, arts, literature, medicine, and general science. This section also provides an overview of the 4 major subsections: identification, screening, eligibility, and data abstraction.

Figure 1. Flow diagram of the proposed search study. WoS: Web of Science.

### Identification

The systematic review process can be divided into 3 main phases, each of which is responsible for selecting several appropriate papers for this report. The first step is the identification of keywords, followed by a search for related and analogous terms using resources such as the thesaurus, dictionaries, encyclopedias, and earlier research. Consequently, after determining all pertinent keywords, search strings for the databases Scopus, Web of Science, and EBSCO (Textbox 1) were formulated. This review successfully retrieved 394 papers from 3 databases during the first step of conducting a systematic review.

**Textbox 1.** Search strings.

<table>
<thead>
<tr>
<th>Database</th>
<th>Search String</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scopus</td>
<td>TITLE-ABS-KEY (&quot;pervasive gam*&quot; AND design* AND (interaction OR interactive)) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013)) AND (LIMIT-TO (PUBSTAGE, &quot;final&quot;)) AND (LIMIT-TO (DOCTYPE, &quot;ar&quot;)) AND (LIMIT-TO (SRCTYPE, &quot;j&quot;))</td>
</tr>
<tr>
<td>Web of Science</td>
<td>ALL= (&quot;pervasive gam*&quot; AND design* AND (interaction OR interactive)) and Articles (Document Types) and 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years)</td>
</tr>
<tr>
<td>EBSCO</td>
<td>ALL= (&quot;pervasive gam*&quot; AND design* AND (interaction OR interactive)) and Articles (Document Types) and 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years)</td>
</tr>
</tbody>
</table>
Screening
During the preliminary screening, duplicate papers were eliminated from consideration. In the first phase, we decided not to include 29 articles. In the second phase, we used several inclusion and exclusion criteria developed to screen 73 articles. The first criterion was the literature, which consisted of research articles, as this is the most important source of useful information. Publications in the form of systematic reviews, reviews, meta-analyses, meta-syntheses, book series, books, chapters, and conference proceedings were excluded in the current investigation. In addition, the review focused solely on papers written in English. It is essential to remember that the plan was to select publications across 10 years, from 2013 to 2022. Thus, 321 publications were disregarded because they did not meet certain criteria.

Eligibility
A total of 44 articles were prepared for the third step, which is referred to as eligibility. At this stage, the titles of all the articles and the primary content they contained were subjected to a comprehensive review to verify that the inclusion requirements had been reviewed and that the articles were suitable additions to this study following the research goals. Consequently, 4 reports were excluded because they did not consist of pure scientific articles based on empirical evidence. Finally, 40 articles were included in the final analysis (Textbox 2).

Textbox 2. Selection criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Language: English</td>
</tr>
<tr>
<td>• Timeline: 2013-2022</td>
</tr>
<tr>
<td>• Literature type: journal (only research articles)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Language: non-English</td>
</tr>
<tr>
<td>• Timeline: &lt;2013</td>
</tr>
<tr>
<td>• Literature type: journal (book chapter and conference proceeding)</td>
</tr>
</tbody>
</table>

Data Abstraction and Analysis
An integrative analysis, which is one of the examination techniques used to analyze and synthesize different research designs, was conducted in this study. This technique was used to integrate qualitative, quantitative, and mixed methods. The primary focus of the specialists’ research was on the formulation of pertinent topics and subtopics. The data collection phase was the initial step in developing the theme. We thoroughly combed through a collection of 40 papers, searching for statements or information that responded to questions raised by this study. The second step involved the expert and us analyzing the situation regarding pervasive games within design thinking, determining meaningful groups to form. The 4 primary themes that emerged from the design and development, interaction and technology, user and experience, and evaluation and service need to be overcome. We reviewed every developed theme from this point onward, including any themes, concepts, or ideas that have any connection. We established themes based on the findings of the study within the framework of this research. In this instance, a log was kept during the data analysis to document any analyses, opinions, puzzles, or other ideas pertinent to data interpretation.

We also compared the findings to resolve any inconsistencies that may have occurred while developing the theme. It is important to note that we discussed with academics any inconsistencies that may have occurred within the themes and tried to resolve the inconsistencies. In conclusion, the developed themes underwent minor adjustments to ensure coherence. Two specialists carried out the examinations: one specialized in the games industry and the other specialized in the game academic areas. This was performed to ensure that the problems were valid. By establishing domain validity, the expert review phase helped ensure that each subtheme was clear, important, and adequate. Adjustments were made at our discretion based on comments and feedback from industry professionals.

Results

Overview
By extending the game experience into the physical domain, pervasive games represent a new era of entertainment and interaction. Pervasive game design is one of the many issues that must be reconsidered. As the development of pervasive games becomes more and more successful, the need to incorporate new design concepts becomes increasingly urgent. There were 40 articles extracted and analyzed based on a searching technique. All articles were classified according to 4 main themes: design and development (11/40, 28%), interaction and technology (15/40, 38%), users and experience (9/40, 23%), and evaluation and service (5/40, 13%; Table 1).
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Journal</th>
<th>Title</th>
<th>Available in the database?</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arango-López et al [14], 2021</td>
<td>Universal Access in the Information Society</td>
<td>GeoPGD: methodology for the design and development of geolocated pervasive games</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Arango-López et al [15], 2021</td>
<td>Universal Access in the Information Society</td>
<td>Using geolympus to create a pervasive game experience in the higher education context</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Costa et al [16], 2021</td>
<td>Multimodal Technologies and Interaction</td>
<td>An Interactive Information System that supports an augmented reality game in the context of game-based learning</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Oliva-Maza et al [17], 2021</td>
<td>IEEE Revista Iberoamericana de Tecnologias del Aprendizaje</td>
<td>Mystery of the runaway letabytes: inclusive assessment of phonological awareness with tangible gamification</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Santos et al [18], 2021</td>
<td>JMIR serious games</td>
<td>Promoting physical activity in Japanese older adults using a social pervasive game: randomized controlled trial</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Alavesa et al [19], 2020</td>
<td>Multimedia Tools &amp; Applications</td>
<td>Embedding virtual environments into the physical world: memorability and co-presence in the context of pervasive location-based games</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bellotti et al [20], 2019</td>
<td>IEEE Transactions on Games</td>
<td>REAL: reality-enhanced applied games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Costa et al [21], 2020</td>
<td>Information</td>
<td>Design of a mobile augmented reality platform with game-based learning purposes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bonillo et al [22], 2019</td>
<td>Multimedia Tools &amp; Applications</td>
<td>Developing pervasive games in interactive spaces: the JUGUEMOS toolkit</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Eckstein et al [23], 2019</td>
<td>Entertainment Computing</td>
<td>Smart substitutional reality: integrating the smart home into virtual reality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Leone [24], 2019</td>
<td>IxD&amp;A (Interaction Design and Architecture(s))</td>
<td>U’Game—a toolkit for urban gaming</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Santos et al [25], 2019</td>
<td>JMIR serious games</td>
<td>Effects of social interaction mechanics in pervasive games on the physical activity levels of older adults: Quasi-Experimental Study</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Santos et al [26], 2019</td>
<td>Journal of Rehabilitation and Assistive Technologies Engineering</td>
<td>Pervasive game design to evaluate social interaction effects on levels of physical activity among older adults</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Guo et al [27], 2018</td>
<td>Information</td>
<td>Ontology-based domain analysis for model driven pervasive game development</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Klaassen et al [28], 2018</td>
<td>Sensors (Basel, Switzerland)</td>
<td>Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Valente et al [29], 2018</td>
<td>Personal &amp; Ubiquitous Computing</td>
<td>A method to assess pervasive qualities in mobile games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kasapakis and Gavalas [30], 2017</td>
<td>Multimedia Tools and Applications</td>
<td>Occlusion handling in outdoors augmented reality games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Maia et al [31], 2017</td>
<td>Entertainment Computing</td>
<td>LAGARTO: A LocAtion based Games AuthoringTOOl enhanced with augmented reality features</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Valente et al [7], 2017</td>
<td>Requirements Engineering</td>
<td>Mapping quality requirements for pervasive mobile games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Study, year</td>
<td>Journal</td>
<td>Title</td>
<td>Available in the database?</td>
<td>Remarks</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>-------</td>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Jong [32], 2016</td>
<td><em>British Journal of Educational Technology</em></td>
<td>Teachers’ concerns about adopting constructivist online game-based learning in formal curriculum teaching: the VI-SOLE&lt;sup&gt;2&lt;/sup&gt; experience</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mora et al [33], 2016</td>
<td><em>Journal of Ambient Intelligence &amp; Smart Environments</em></td>
<td>From interactive surfaces to interactive game pieces in hybrid board games</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sekhavat [34], 2016</td>
<td><em>International Journal of Computer Games Technology</em></td>
<td>KioskAR: an augmented reality game as a new business model to present artworks</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Guo et al [35], 2015</td>
<td><em>International Journal of Multimedia and Ubiquitous Engineering</em></td>
<td>RealCoins: A case study of enhanced model driven development for pervasive games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kasapakis and Gavalas D [36], 2015</td>
<td><em>Journal of Network &amp; Computer Applications</em></td>
<td>Pervasive gaming: status, trends and design principles</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kasapakis et al [5], 2015</td>
<td><em>Personal and Ubiquitous Computing</em></td>
<td>Pervasive games field trials: recruitment of eligible participants through preliminary game phases</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lemcke et al [37], 2015</td>
<td><em>Personal and Ubiquitous Computing</em></td>
<td>RouteMe: a multilevel pervasive game on mobile ad hoc routing</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lv et al [38], 2015</td>
<td><em>Personal &amp; Ubiquitous Computing</em></td>
<td>Touchless interactive augmented reality game on vision-based wearable device</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Plöhn et al [39], 2015</td>
<td><em>Electronic Journal of e-Learning</em></td>
<td>Dynamic pervasive storytelling in long lasting learning games</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>El-Nasr et al [40], 2015</td>
<td><em>Entertainment Computing</em></td>
<td>Data-Driven Retrospective Interviewing (DDRI): A proposed methodology for formative evaluation of pervasive games</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Sra and Schmandt [41], 2015</td>
<td><em>Personal &amp; Ubiquitous Computing</em></td>
<td>Expanding social mobile games beyond the device screen</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Evans et al [42], 2014</td>
<td><em>Personal &amp; Ubiquitous Computing</em></td>
<td>The Malthusian Paradox: performance in an alternate reality game</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Maggiorini et al [43], 2014</td>
<td><em>Multimedia Systems</em></td>
<td>Opportunistic mobile games using public transportation systems: a deployability study</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Schmitz et al [44], 2014</td>
<td><em>Pervasive and Mobile Computing</em></td>
<td>The impact of coupled games on the learning experience of learners at risk: an empirical study</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Chamberlain et al [45], 2013</td>
<td><em>Entertainment Computing</em></td>
<td>Them and Us: an indoor pervasive gaming experience</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Chen et al [46], 2013</td>
<td><em>International Journal of Human-Computer Interaction</em></td>
<td>Your way your missions: A location-aware pervasive game exploiting the routes of players</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Coenen et al [47], 2013</td>
<td><em>Journal on Computing and Cultural Heritage</em></td>
<td>MuseUs: case study of a pervasive cultural heritage serious game</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kristiansen [48], 2013</td>
<td><em>International Journal of Mobile Human Computer Interaction</em></td>
<td>Design games for in situ design</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Neustaedter et al [49], 2013</td>
<td><em>Personal &amp; Ubiquitous Computing</em></td>
<td>Creating scalable location-based games: lessons from Geocaching</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Papagiannakis et al [50], 2013</td>
<td><em>Journal of Universal Computer Science</em></td>
<td>A multimodal ambient intelligence environment for playful learning</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Design and Development

Development Mode

Pervasive games are a completely new game form that extends game experiences into the physical world by incorporating information and communication technologies into the fabric of players’ real-world environments. This emerging game mindset makes it difficult for developers to explore technologies and methods for providing a high-quality interactive experience for users and designing novel and compelling forms of content [36]. Model-driven development and domain-specific modeling have succeeded in many open or in-house information system scenarios. However, its application to games is rare and immature. Guo et al [27] illustrated solutions by describing a pervasive game case’s technical and procedural development, incorporating model-driven development into the game development process. Furthermore, Guo et al [27] proposed a lightweight domain analysis that can be incorporated into the pervasive game development process. The domain analysis process is based on a game ontology that provides game design and domain analysis. They introduced the ontology, demonstrated its application in the domain analysis procedure, and evaluated and discussed the quality of the ontology using a user acceptance survey. The test results indicated that most potential users found the ontology useful and simple to use.

Development Tool

When developing pervasive games, programmers must deal with high turnover, cheating, and heterogeneity of devices and platforms. LAGARTO is an authoring tool for constructing location-based mobile games with augmented reality (AR) capabilities. These games are a subclass of pervasive games in which the game’s progression depends on the location of the players. Maia et al [31] gleaned requirements for designing the LAGARTO software architecture from a literature review and focus group meetings. The tool consists of a web-based application for creating and managing games and a mobile app for tracking players. Their primary objective is to develop software that enables nonprogrammers to design, construct, and run location-based mobile games. The development of pervasive games is slowing down because of the multiple challenges that these games present to their creators, such as the vast array of interaction paradigms and the complexity of developing applications where so many innovative technologies converge. Bonillo et al [22] introduced JUGUEMOS, a toolkit designed to assist developers in creating pervasive games for interactive spaces. The toolkit addresses 3 difficulties that arise during the development of pervasive games: integrating heterogeneous devices, managing multiple displays, and facilitating game coding. They conducted case studies exploring the toolkit’s expressivity, its ability to support collaborative multidisciplinary experiences, and its potential to support interactive experiences outside our interactive space. With the addition of pervasive or narrative components, however, the complexity of construction and support increases, necessitating a tool to manage the information appropriately and dynamically. Space and the pervasiveness of social interaction collaborate to achieve objectives by exchanging data between multiple pervasive games to improve the game experience. Arango-López et al [15] described a platform (Geolympus) that enables the creation of game experiences based on the player’s location.

Development Direction

In recent years, games’ design and development incorporating web-based elements into real environments have expanded. Such games include pervasive games, which seek to enrich the game world by combining these 2 realities to immerse the player more deeply in the story. The implementation of pervasive games has increased player engagement and motivation. New technologies have enabled advancements in the design and development of video games. Pervasive games have not been an exception, and combining traditional game elements with real-world elements and scenarios has made it possible to enhance motivation and user experience. However, previous research has demonstrated the need for a procedure to guide the design and development of pervasive games. In light of this, Arango-López et al [14] developed GeoPGD, a methodology that integrates the design of geolocated narrative as the core of the game experience. This methodology guides designers and developers through the various stages of creating pervasive games and provides tools for defining narrative elements, locations, and user-pervasive game interactions. Furthermore, Płonś et al [39] described the Dynamic Pervasive Storytelling (DPS) approach and the design of the pervasive game Nuclear Mayhem, which was created to support a course in web game development at Nord-Trndelag University College in Norway. Nuclear Mayhem runs alongside the course for 9 weeks and requires specific elements to become part of the players’ everyday lives, reminding them of the game they are playing. The game DPS as a model aims to increase the universality of the game and support the player’s ongoing level of awareness in the game by incorporating real-world events. DPS uses real-world events as a building block to create the entire game story before the game begins by incorporating current events into its design and increasing the universality of the experience and in-game awareness as the game unfolds. This paper concludes that DPS is a promising approach to developing game narratives because it increases the universality of the game and promotes in-game awareness.

Mobile culture has given rise to many context-based products such as location- and hashtag-based apps. This poses new
challenges for designers. Design methodologies that recognize context and allow it to influence design concepts are required. Kristiansen [48] focused on a design problem where live design practices can enable the early design process: designing a pervasive game. Pervasive games are computer games that use cities as playgrounds and cell phones with GPS capabilities. Some contextual design approaches already exist, but the authors propose an approach that requires designers to conceptualize and implement ideas in the field, that is, at the game’s location. The challenge is to develop a creative approach that includes on-site design work and generates game concepts for pervasive games. The proposed design approach, called site storming, was based on games involving individual scenario exploration of the site and various types of game cards, followed by a collective evaluation of the ideas generated. A series of evaluations showed that designers enjoyed using this method, stimulated idea generation, and live design influenced their design concepts.

The pervasive game aims to take computer games out of the PC and into the “real world” of cities, streets, and parks. This real-world physicality makes for an enjoyable player experience but presents a unique challenge for designing and coordinating such games. Neustaetter et al [49] explored how location-based games can be designed to overcome this scalability challenge—using Geocaching to determine how it has maintained user engagement and growth over the past decade. The findings show that Geocaching benefits from direct player creation of simple and complex game content. Geocaching also simplifies game choreography by allowing players to monitor game content, other players, and even nonplayers. Recently, developments in virtual reality technology have allowed users to consume immersive content from the comfort of their living rooms. Alternate reality promises to enhance this experience by incorporating physical environments into the simulation. Smart substitute reality (SSR) is a novel approach that combines the passive haptics of substitute reality with the active capabilities of a smart home environment. SSR is intended to serve as the basis for serious games. Eckstein et al [23] discussed the concept of SSR while implementing a prototype in our Smart Lab. They created multiple web-based environments with varying degrees of mismatch to the real world and added selected objects to induce additional tactile and thermal stimuli to enhance immersion.

Interaction and Technology

Outdoor Location

With the rise of pervasive computing, location-aware pervasive games are increasing rapidly. Chen et al [46] presented the design and implementation of Your Way Your Missions. This location-aware pervasive game uses the player’s route to combat the inefficiencies of location- and radius-based information adaptation. Your Way Your Missions provides the player with a Google Maps–based tool to predefine the route and relies on self-reporting to obtain the player’s planned route. Such a design returns the task to the player based on the task’s location attributes and the player’s intended route. The results show that the route predefinition and self-reporting approach is an effective way to obtain the player’s planned route. This adaptation of information based on the player’s planned route can help them locate the task at any time and location. Leone [24] also described this location-based outdoor interaction in his article. He described this small community’s view of the connection between the real world and the video game world using physical and digital support. The paper highlights 3 games produced through participatory processes and argues for the role of urban and popular games as tools to promote citizenship. Other scholars have also suggested that using augmented or web-based field trips and pervasive educational games can potentially transfer learning content to the real world outside the classroom. Lemcke et al [37] presented the pervasive educational game RouteMe, which contextualizes the rather abstract topic of routing in ad hoc networks. The game was designed for college-level courses as a motivational supplement to these courses to enhance the learning experience. Students take on the role of routing nodes or applications with routing requirements. In 3 successive levels of difficulty, they are introduced to game concepts, learn basic routing mechanics, and become aware of the general limitations and capabilities of routing nodes.

With the increasing popularity of personal communication devices, the demand for location- and environment-based mobile services is growing exponentially. The mobile game as a service is no exception. Unfortunately, unlike other services, location- and context-based games strictly require near-field communication to create teams and arenas with nearby players. As currently used technologies have scalability (Bluetooth) or energy (Wi-Fi) limitations, opportunistic networks are considered a viable solution to engage large numbers of players in a larger area. However, it remains unclear how the increased latency and probabilistic message forwarding introduced by opportunistic networks will affect the player experience. Maggiorini et al [43] addressed these issues by simulating the deployment of a connection-based game on opportunistic networks provided by public transportation systems in 3 cities: Milan (Italy), Edmonton (Alberta, Canada), and Chicago (Illinois, United States). Moreover, they studied an opportunistic cooperative version of a well-known independent game for playability and scalability reasons. The emphasis on this particular game promotes the use of public transportation systems. Emerging pervasive games use sensors, graphics, and web technologies to provide immersive gameplay interactions integrated with the real world. Existing pervasive games typically rely on the device screen to provide game-related information and ignore opportunities to include new types of contextual interactions such as jumping, punching gestures, or even voice as game input. Sra and Schmandt [41] presented the design of Spellbound, a team-based physical mobile game, to help us understand how to design pervasive games that foster a spirit of togetherness. Sra and Schmandt [41] also briefly discuss how solidarity and playfulness transform physical movements into an ideal activity. Spellbound is an outdoor, team-based physical game. It takes advantage of the aforementioned opportunities to combine real-world movements such as jumping and spinning with the web-based world. It also replaces touch-based input with voice interaction. It uses custom hardware to provide tactile feedback, embodying the true spirit of social play that characterizes traditional children’s games.
They believe that Spellbound is a form of digital outdoor play anchoring enjoyment in physical movement, social interaction, and tangible feedback.

Industry Application

Pervasive games have also contributed in terms of industry domains. Coenen et al. [47] presented a case study of MuseUs, a popular pervasive game for museums that operates as a mobile phone app. During the museum visit, players are encouraged to create their exhibits and are guided through the process by application. The aim was to provide a learning effect during a museum exhibit visit. The core of the MuseUs experience is that it does not require a predetermined path through the museum and does not interfere with the exhibit itself. In addition, the application encourages visitors to see elements of cultural heritage in a new light, enabling a personal narrative while building a personal exhibition. Moreover, Sekhavat [34] described the architecture of KioskAR, a pervasive AR game. This game introduces a new business model that enables players to use AR to display their artworks in a web-based kiosk while enjoying the game. In addition to the competition, this game requires social interaction between players to earn more points. A user study evaluated the presence and usability of the application. The experiment showed that KioskAR could achieve a high level of usability and presence.

Interaction Mode

Creating pervasive games using emerging interactive technologies is becoming increasingly popular. Lv et al. [38] designed and evaluated a touchless motion interaction technique for developing interactive and AR games on vision-based wearable devices. Users interact with AR games by making dynamic hand or foot gestures in front of the camera, which triggers interactive events with objects in the scene. As evidence, 3 elementary AR games with 11 dynamic gestures were created using the proposed touchless interaction technique. Finally, they presented a comparative evaluation to demonstrate the social acceptability and usability of the touchless approach running on a hybrid wearable frame or Google Glass and assess workload, user sentiment, and satisfaction. Recently, developments in interactive surfaces and tangible user interfaces offer new opportunities for pervasive games that combine traditional board games’ social affordances with the interactivity of video games. Mora et al. [33] proposed a strategy centered on tokens, constraints, spatial representations, and interactive events in this research area. Rather than using interactive surfaces as the primary interaction medium, this approach relies on physically manipulating interactive, computer-enhanced game components on standard surfaces. Furthermore, AR in mobile game development, where computer-generated graphics are used to enhance the player’s view of the physical world, is becoming increasingly common.

AR Overview

In AR applications, it is common for objects to be partially or completely obscured by physical objects; if not handled properly, the visualization of obscured objects can often mislead the user’s perception. Kasapakis and Gavalas [30] presented 3 alternative geolocative raycasting techniques designed to help outdoor pervasive game developers solve the occlusion problem by integrating real-time building recognition to generate a realistic field of view for the player. The geographic raycasting approach has been implemented in the location-based pervasive game Eliminate Order, which uses publicly available free building data to calculate the player’s field of view in real time. The proposed algorithm applies to many sensor-based AR applications and can be ported to any real-world environment. A few years earlier, Evans et al. [42] examined the design of the Malthusian Paradox and highlighted how it redefines the framework of performance in sometimes disturbing ways by blurring the traditional roles of performer and audience. Players engage with the game through various interactive channels, including performative group performances and one-on-one interactions with game characters in public settings, using low-end and high-tech physical and web-based artifacts such as customization and third-party websites. Players and game characters communicate in planned and unplanned ways via phone and social media. We reflect on the production and choreography of the game, including the dynamic nature of strong episodic narratives driven by professionally produced short films that attempt to respond to the player’s actions and the difficulties of designing for engagement across hybrid and time-dilated performance spaces.

Realistic 3D environments, such as existing city models, have the potential to serve as a bridge between the physical and the environments in pervasive games. Smooth attention shifts and transitions between these 2 realities are largely unexplored in the context of pervasive games. Alavesa et al. [19] conducted 2 field experiments using a pervasive live-action role-playing game to investigate the effects of transitions between the web-based and physical environments on coexistence and memorability. Although there were few differences in copresence during play, they highlighted the subtleties in the social structure of the universal game. 3D environments are more memorable because of their spatial similarities to the physical world. We then identified 2 important factors to consider when integrating web-based environments into pervasive games: the structuring of social interactions and spatial realism. The emerging genre of pervasive games combines reality with computation. Bellotti et al. [20] explored the concept of “reality-enhanced games” through a simple set of pervasive games. A reality-enhanced game is a model that links game mechanics to the outcomes or measurements of real-world activities. The prototype game was created as part of the TEAM Industrial Research Project, which aims to develop mobile apps for flexible and collaborative mobility. The proposed game is an example of various user interfaces that are believed to be useful for various important scenarios, goals, and user types, particularly for improving driving styles. This concept allows pervasive game developers to focus on their game logic while seamlessly leveraging a variety of in-field sensors, and it is generic.

AR is an emerging technology that superimposes objects onto the physical world. Recent computer and mobile technology advances have made AR increasingly accessible in education. Costa et al. [21] described an educational mobile AR platform. The platform includes a mobile app consisting of a pervasive

https://games.jmir.org/2023/1/e42878

JMIR Serious Games 2023 | vol. 11 | e42878 | p.647

(page number not for citation purposes)
game designed to encourage learning about the universe. In addition, it includes a backend that enables teachers to present information about celestial bodies and create multiple-choice questions to assess students’ knowledge of the participants they are studying. The mobile app provides users with real-world physical movement and social interaction while playing the game; therefore, it falls within the paradigm of pervasive games. In addition to engaging students to play games, they believe that this platform can be implemented in both informal and formal learning environments. Mobile AR applications are gaining prominence in education, but there is a need to design educational games that are both appropriate and enjoyable. Costa et al [16] described an interactive information system to facilitate the implementation of a pervasive game for game-based learning. PlanetarySystemGO includes a location-based mobile AR game that facilitates learning about cosmic objects and planetary systems and a web application that interacts with mobile device apps. This resource can be used for web-based and face-to-face classes, which are useful in socially isolated situations, such as those caused by the COVID-19 pandemic. In addition, adding web applications with a backend to the information system allows for the inclusion of curriculum content based on the student’s grade level. Similarly, teachers are expected to use the information system to incorporate content that they feel is appropriate for their teaching grade level. Therefore, it is critical to provide them with the professional development they need to use this resource. Teachers found this resource helpful in motivating students to learn, recognizing that the web application facilitated the introduction of appropriate curriculum content and was useful in assessing student performance in games.

**Users and Experience**

**Children’s User Experience**

There is a growing body of research on pervasive outdoor games, with most studies focusing on adult mobile phone players. Most published assessments of player experience in such games are based on anecdotal descriptions and postgame surveys. The latter approach is particularly difficult to implement when the game testers are children. Until now, observations of games have been ad hoc and based on unstructured observations, making it difficult to draw reliable conclusions from observations and compare games. The Outdoor Play Observation Scheme and GroupSorter are 2 methods developed specifically to assess the player experience of children’s outdoor play [51]. They discussed their application in 3 case studies and concluded that Outdoor Play Observation Scheme is useful for quantifying various play behaviors in outdoor play. GroupSorter adds qualitative data about play experiences. Moreover, the application of GroupSorter is not limited to game development; it can be used to collect user input in other contexts. In the same year, Papagiannakis et al [50] described the design, development, and evaluation of AmI Playfield, a technical framework for learning applications designed to create challenging learning conditions through play. An AmI Playfield is a guided ambient intelligence. It emphasizes using kinesthetic and collaborative technologies in a natural, play-based learning environment. It also incorporates performance measurement techniques. The Apple Hunt application was created to test and evaluate AmI Playfield, engaging child learners in arithmetic thinking through kinesthetic and collaborative play. Simultaneously, AmI technology observes behind the scenes. Apple Hunt has been assessed using a combination of methods suitable for young testers, and the Children’s Commission is presented as a promising assessment method for children. The results indicate the system’s high capacity to stimulate thinking and enjoyment, thanks to the learners’ whole-body kinesthetic play and teamwork.

Acquiring literacy skills requires developing phonological awareness, a metacognitive skill. This allows thinking about and manipulating language structures from an early age. Learning requires multisensory stimulation as well as curiosity and interest. Communication technologies play an important role in this, as learning-oriented use can stimulate interest and thinking in current digital natives. Oliva-Maza et al [17] described a pervasive game with tangible technology that can be used for interventions and diagnostic, formative, and summative assessments. Multiple case studies were conducted with 4 kindergarteners over 2 weeks to test the user experience. They were given a multimedia pervasive adventure game in which they had to use tangible technology to solve phonological awareness challenges to reach the final goal. Observations were recorded in an evaluation model for posttriangulation data analysis and subsequent qualitative analysis of pre- and postintervention data. The results showed improvements in phonological awareness and traces of gamified experiences. These encouraging results justify the application of these methodologies to a learner-centered education model.

**Adults’ User Experience**

The emergence of pervasive technologies has sparked an interest in designing and developing pervasive games. There are numerous designs of pervasive games for adult users. Chamberlain et al [45] presented Them and Us, a pervasive indoor game that uses dramatic processes to encourage social interaction within the context of the game. The Them and Us gameplay brings a group of people together in a space to interact with each other. At the same time, location-based technology reveals us the locational nature of the social interactions occurring in the space (who, where, and when). It allows us to score based on these interactions. Them and Us took a narrative-based approach in which a script instructed participants on how to play the game and how their social behaviors were incorporated into the game mechanics. This new interactive game-based artwork blends the player’s physical and web-based worlds to create a new and exciting experience for them. Designed specifically for adults, the pervasive game—Them and Us—emphasizes the use of audience and performer tracking as a means of encouraging interaction within the play space.

The pervasive game has also made a significant contribution to the education of adults. Providing pervasive game-based learning scenarios for at-risk students is effective and motivating. Schmitz et al [44] provided a detailed example of an educational environment in which a mobile game is combined with a PC browser game. They evaluated how this pairing helps the target group engage and learn. The study surveyed 19 people, ranging in age from 17 to 21 years, who participated in the game.
and explored it. The results of the 7-week game study suggest that pervasive games can increase learners’ interest in a topic while also supporting learning activities. Other studies have been conducted in the context of intense debate about how pervasive games can be used to provide new constructivist learning opportunities for students. Jong [32] developed a theoretical foundation for the Virtual Interactive Student-Oriented Learning Environment (VISOLE). It was a pedagogical framework for implementing constructive web-based gamified learning in teaching formal curricula. In addition to highlighting the pedagogical features of VISOLE, Jong [32] discussed the Stages of Concern model, which investigates 118 teachers’ concerns about this educational innovation in terms of 5 categorical concerns: evaluation, information, management, consequences, and refocusing. All participants had completed introductory VISOLE training before taking the Stages of Concern survey. He was also able to develop more precise interventions tailored to the actual needs of teachers to help him use this diagnostic knowledge to implement VISOLE in their schools. The findings also shed light on how to design, develop, or adapt pervasive games to aid learning and teaching in the classroom.

**Older Adults’ User Experience**

Promoting active lifestyles among older adults can have a significant positive impact on their quality of life. Pervasive games aim to create more fun and engaging experience by incorporating real-world elements into the game. The innovative mechanics of pervasive games, combining real and web-based worlds, can further engage and motivate older adults to achieve this goal. They can be a powerful strategy to encourage physical activity among older adults as they are integrated into players’ lives and naturally promote more casual play. Santos et al [26] designed and evaluated the feasibility of a pervasive game using social interaction as a case study to understand how game design elements can influence the physical activity levels of older adults. They created a mobile, location-based pervasive game and tested its feasibility as an experimental system among community-dwelling older volunteers in Kyoto, Japan. Participants found the theme and visual style of the game appropriate, and the rules and objectives of the game were easy to understand. The game was praised for its challenging and entertaining nature. Further research suggests that future iterations of the system will require special attention to the complexity of the controls and new ways to connect players when few people are playing or when they are too far apart.

In the same year, the goal of Santos et al [25] was to test the impact of changing a specific design element of a popular game for older adults, namely social interaction, on physical activity levels. Therefore, they experimented by comparing 2 variations of the same popular game over 4 weeks: the test group received social interaction, whereas the control group received no social interaction. In both versions, players had to walk to physical locations to collect cards, but the socially interactive version allowed people to cooperate in obtaining more cards. The impact on each group was assessed using the number of steps taken per week, and the number of places visited was used to measure game activity. A total of 32 people were recruited for the experiment (no social interaction=15; social interaction=17); 18 people persisted until the end (no social interaction=7; social interaction=11). The experiment found that the social interaction design element of the popularity game may have had some positive effects on promoting physical activity. Subsequently, in 2021, Santos et al [18] continued with the same study, comparing 2 variations of the same pervasive game over 4 weeks with the same purpose and type of experiment: the test group received social interaction whereas the control group did not. In this test, 20 participants were recruited for this experiment (no social interaction group, n=10; social interaction group, n=10), 18 of whom remained active until the end of the study (no social interaction group, n=9). After several tests of pervasive games with older adults, they concluded that the social interaction design element of pervasive games might positively affect physical activity. Although other factors may have influenced this effect, it has tremendous implications for the fitness and health of older adults.

**Evaluation and Service**

**Evaluation in the Area of Health**

Recently, games have gained attention as a medium for changing health behaviors. However, current projects on healthy games face a major obstacle: the evaluation methods used to assess and formally evaluate such games do not adequately measure the acceptability and integration of the games into participants’ lives. These are important components of ensuring participants adhere to the games to induce behavior change. El-Nasr et al [40] described the data-driven retrospective interviewing method that we developed for formative evaluations. Used in a naturalistic setting, the data-driven retrospective interviewing examines how participants embrace the game and integrate it into their lives. The procedure consists of several steps. First, a game is designed to collect behavioral data, which are then analyzed to provide a basis for developing interview questions that retrospectively understand and reflect participants’ behavior. The contribution of this paper lies in its assessment methodology and application to SpaPlay, a universal health game [52]. As a result of this investigation, designers could rethink and improve their game design as they discovered that users were not using the game as initially envisioned. Such results would not have been possible with traditional evaluation techniques.

Self-monitoring, goal setting and coaching, education, and social support are daily care strategies for patients with chronic diseases. Various tools, such as mobile digital coaching systems linked to wearable sensors, pervasive games, and patient portals for personal health records, have been developed to assist patients with chronic conditions and their caregivers in achieving the ideal of self-management. Klaassen et al [28] described a platform that integrates these tools to help young patients self-manage their diabetes through educational games, monitoring, and motivational feedback. The platform’s design references health care, persuasive system design, and serious game design principles. A coach is a game guide that provides personalized feedback on the user’s daily care-related activities and helps them progress in the game world. User evaluations with patients under pediatric supervision suggest that using mobile technology in conjunction with web-based elements is feasible. However, assumptions about how users connect to the
platform are not met in practice, resulting in a suboptimal user experience. They discuss the challenges and provide recommendations for advancing the integration of coaching and gamification platforms prevalent in medical practices.

**Evaluation in Other Areas**

Some scholars have explored a new approach to conducting user evaluation trials for pervasive games. In a case study, Kasapakis et al [5] examined the evaluation process of the pervasive role-playing game *Barbaros* consisting of preliminary and main implementation phases. The former is freely accessible to anyone and can be played at any time or place without the investigator’s team’s organizational and coordination investment. The latter defines 3 interdependent player roles that must cooperate in the scavenger hunt to achieve a common game goal. Players’ eligibility to participate in the main game phase is determined by their position in the previous phase. Using concepts from cultural theory, they designed the preparation phase as a dynamic environment from which potential evaluators would emerge. The main hypothesis being investigated is that implementing such a cost-effective preparation phase could recruit highly qualified participants for user trials of the pervasive game study prototype, thereby increasing the reliability and quality of evaluation results.

The requirement’s engineering community has not yet given enough attention to games. It becomes even more critical as we move toward newer forms of games, such as pervasive games. Pervasiveness (the characteristic that distinguishes traditional digital games) can mean being everywhere, pervading something, or propagating something through physical space. Each time a game is played, the boundaries of the game expand from the fictional world to the real world, indicating pervasiveness. Pervasive games, also known as alternate reality games, transmedia games, and cross-media games, are a new type of digital entertainment that has developed in various forms. The rapid development of this new type of digital product can be attributed to sensor technologies, networking capabilities, AR systems, computer vision technologies, the internet, and, in particular, mobile devices. Valente et al [7] defined it as pervasive games played on mobile devices. Mobile devices are currently the main drivers for the promise of pervasive games. Using various information sources, they compiled a list of interrelated characteristics critical to the success of these games. They presented a 2-level conceptual map of nonfunctional requirements based on these fundamentals to help achieve pervasiveness in pervasive mobile games.

Furthermore, Valente et al [29] proposed a new method for evaluating the quality of pervasiveness in pervasive games that can be adapted and applied to other pervasive applications. This approach produces a quality report consisting of a quality spreadsheet (containing metric values and comments) and a quality vector (representing the game quality profile in the form of a histogram). In addition, they can compare quality vectors based on similarity criteria. They applied the proposed approach to commercial and academic prototype games to elucidate their universality characteristics and identify ways to improve the overall quality of these games. In this sense, what distinguishes them from traditional digital games is the universality.

**Discussion**

**Principal Findings**

The analysis and results of the systematic literature review broadly divided these articles into 4 directions. These 4 directions are the direction of design and development, the direction of interaction and technology, the direction of the user and experience, and the direction of evaluation and service. These 4 directions appear to be independent of each other, but they are interdependent, which is consistent with the idea of this study. At this point, these directions can be correlated to the research objectives of this study one by one. Therefore, after abstracting and simplifying the themes of these literature analysis results, it is possible to sublimate them into corresponding design fields. Namely, the direction of design and development corresponds to the field of game design; the direction of interaction and technology corresponds to the field of interaction design; the direction of user and experience corresponds to the field of experience design; and the direction of evaluation and service corresponds to the field of service design. Therefore, these 4 design fields are the 4 research objectives to be analyzed in this study: game design, interaction design, experience design, and service design. They are also important indicators for analyzing the pervasive game from the perspective of different design thinking.

The game design aspect of pervasive games presents polymorphic research, which can be roughly divided into 3 categories: the development mode of game design, the development tool of game design, and the development direction of game design. Because pervasive games are a class of games, they naturally have a game design and development properties. As early as 2015, Kasapakis and Gavalas [36] introduced the historical development lineage of pervasive games, player acceptance factors, game development principles, and future development trends by analyzing 2 generations of 18 successful pervasive games. Subsequently, scholars have proposed different toolkits for developing pervasive games, enriching the toolbox of pervasive game design [15,22,31]. Moreover, the development direction of pervasive game design has never stopped research from 2013 to 2021. Inside these ideas are those that let players create game content directly [49], those that let designers find inspiration in the field [48], and those that let the game itself incorporate smart homes [23]. Furthermore, some research studies pervasive games from a storytelling and narrative perspective. This approach guides designers and developers through the different stages of game development, providing tools to define narrative elements, locations, and interactions between users and the game [15,39]. It is evident from looking at various aspects of pervasive game design, a popular focus of scholars’ research, showing a promising trend of multifaceted and multidimensional development.

The interaction design aspect of pervasive games has been the most studied by scholars, comprising 15 articles. It is only natural that this is the case, as pervasive games are a large category of games, and players must have a variety of interaction designs or interaction technologies in them when they play. First, pervasive games are location-based games, so there is a
large body of scholarship on how they use interactive technology to interact outdoors [24,37,41,43,46]. Second, there were scholars in industrial applications to study how pervasive games can increase interactive communication of artworks and how to improve interactive communication of cultural heritage in museums [34,47]. Touchless interaction is then used in designing the interactive interface and physical user interface of pervasive games [33,38]. Finally, scholars have mostly studied pervasive games combined with AR [16,19-21,30,42]. Furthermore, it can be seen that the technological development factor in the interaction design of pervasive games is influenced. Owing to the limitations of mobile technology in the early years, the interaction design of pervasive games was most explored by SMS text messaging and Bluetooth [42]. By 2021, scholars were exploring ways to integrate AR into the interaction design of pervasive games [16]. Nevertheless, these studies continue to expand the boundaries of interaction design for pervasive games.

The design of the pervasive game experience involves gamers of all ages. Among these articles, pervasive games are designed for children mainly to exercise their reaction and thinking skills [50,51]. Subsequently, there are pervasive games designed for adults, teachers, and students to learn [32,44]. At the same time, there are pervasive games for older adults, indicating some positive effects on promoting physical activity [18]. From the literature, it can see some hints of the research tendencies of scholars: in the early days, they studied the pervasive game design experiences of children users, followed by the pervasive game design experiences of adult users, and in recent years, they focused on the pervasive game design experiences of older adult users. In particular, Santos, a Japanese scholar, and his research team have been studying the effects of social interaction mechanics in pervasive games on the physical activity levels of older adults in recent years [25]. They used multiple experimental measurements to thoroughly analyze older users’ perceptions of a pervasive game designed specifically for them, ranging from the game theme and visual style to game complexity and sociality. The direction of Santos’s research is very similar to the direction of this study, which is to investigate the effects of pervasive games on fitness among older adults, except that the players’ countries are different, the methods used are different, and the details of the specific participants studied are different. However, the research topics are very similar in terms of the effects of pervasive games on fitness among older adults. Nevertheless, the research conducted by Santos et al [25] is of great reference value to this study.

There are also some studies on service design for pervasive games, mainly from the evaluation perspective. This is probably because the service design concept is relatively new and has gradually emerged in design thinking and academic research in recent years. Perhaps it is therefore that the academic research perspective on pervasive games is still limited to simply evaluating the game itself. However, while evaluating the pervasive game itself, scholars are also exploring ways to improve and upgrade the pervasive game under study, which can be seen as an important manifestation of the service design concept. For example, refining the details of game mechanics and upgrading the details of game experience can be considered as research manifestations of service design for pervasive games. As in the case of El-Nasr et al [40] and Klaassen et al [28], the pervasive game for health programs was evaluated and measured for acceptability and integration into the participants’ lives, and at the same time, the pervasive game they studied was optimized and iterated. It is the best example of the service design philosophy running through the pervasive game. Of course, other studies also included such early concepts of service design.

**Limitations**

The first limitation of this study is the database. This study only targeted 3 mainstream literature databases and did not include other databases, such as gray literature or PubMed. This may foreshadow the fact that, although the results obtained in this study through a systematic literature review method are authoritative, they could be more extensive and comprehensive. Another limitation of the study was that the search was conducted up to April 2022, and no articles were found in 2022 at the time of conducting the search, leaving an impression that the search left out publications of 2022.

**Future Studies**

After a systematic analysis, it was found that none of the themes in this literature were related to branding design. However, the article by Kasapakis and Gavalas [36] on the pervasive game’s status, trends, and design principles can be considered a review of the existing successful pervasive games as branding. Plahn et al [39] studied a dynamic and popular narrative approach to creating stories that can reinforce a sense of game maintenance and the brand of pervasive games. Valente et al [7] emphasized the need for nonfunctionality and highlighted aspects, such as cultural awareness, which can be seen as designing a brand strategy for pervasive games. Therefore, this literature does not explicitly go into the brand design of pervasive games but only touches on it to a greater or lesser extent. These results and findings were not coincident with the research object of brand design in this study. Thus, we believe that this systematic literature review has identified a research gap for branding design research on pervasive games and that it could be a future study direction.

**Conclusions**

A new generation of applications that blend the real and digital worlds, such as pervasive games, presents daunting obstacles for software developers. This type of game has recently become a worldwide phenomenon, with thousands of people using their mobile phones to interact with the surrounding physical environment while walking down the street.

At the beginning of the Introduction section, an accurate description of the current situation of the game industry is provided, explaining the entertainment nature and economic value of the pervasive game. In the Methods section, this study adopted a rigorous systematic literature review research process with graphs and tables. This study identified 4 themes: design and development, interaction and technology, user and experience, and evaluation and service. Each of these 4 themes was subdivided into subsections and interpreted. Finally, in the Discussion section, the 4 themes are explained in terms of this study’s objectives: game design, interaction design, experience

https://games.jmir.org/2023/1/e42878

JMI Serious Games 2023 | vol. 11 | e42878 | p.651

(page number not for citation purposes)
Kai et al

design, and service design. It also provides ideas for future research directions in pervasive games based on brand design.

Conflicts of Interest

None declared.

References

Abbreviations

AR: augmented reality
DPS: Dynamic Pervasive Storytelling
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SSR: smart substitute reality
VISOLE: Virtual Interactive Student-Oriented Learning Environment
Gamification and Soft Skills Assessment in the Development of a Serious Game: Design and Feasibility Pilot Study

Luca Altomari*, BEng; Natalia Altomari†, PhD; Gianpaolo Iazzolino*, PhD

1Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende, Italy
2Department of Mathematics and Computer Science, University of Calabria, Rende, Italy
*all authors contributed equally

Corresponding Author:
Luca Altomari, BEng
Department of Mechanical, Energy and Management Engineering
University of Calabria
Ponte Pietro Bucci
Rende, 87036
Italy
Phone: 39 3932671847
Email: luca.altomari@gmail.com

Abstract

Background: The advent of new technologies has had a profound impact on the labor market, transforming the way we work and interact with each other. With the rise of digital tools and platforms, gamification has emerged as a powerful technique for enhancing productivity and engagement in various fields, including human resource management. In particular, gamification has been found to be effective in developing and assessing soft skills, which play a critical role in determining the success of individuals, teams, and organizations.

Objective: We present a serious game that identifies the most sought-after skills in the job market and offers feedback, and we provide a set of guidelines for the creation of serious games.

Methods: We present the serious game Among the Office Criticality (AOC). The AOC game structure involves a set of sequence analysis techniques, which is known as process mining.

Results: The pilot study findings indicate that the game is both engaging and beneficial to subjects, suggesting that the results align with current theoretical perspectives. Furthermore, the study suggests that the obtained data can be extended to the broader population.

Conclusions: This study illustrates a serious game structured according to the needs of the labor market and developed to put the user at the center, using evaluation techniques consistent with the literature, with the aim of constituting an interdisciplinary approach suitable for adequately assessing users and creating value for them.

(JMIR Serious Games 2023;11:e45436) doi:10.2196/45436

KEYWORDS

gamification; soft skills; recruitment; serious games; assessment; process mining; work-life skills

Introduction

Background
The field of human resources has transformed traditional and outdated personnel selection methods through technological advancements, paving the way for more technologically advanced methods that reduce hiring costs and optimize the evaluation of candidates during the recruitment process. This is particularly relevant when it comes to millennials and Generation Z. Gamification is among these methods. It can be defined as “a package of services in which a basic service is enhanced by rule-based service systems that provide the user with feedback and interaction mechanisms that aim to facilitate and support users’ global value creation” [1], and in recent years, it has become an increasingly popular trend in nongaming contexts [2] to increase motivation and engagement [3]. Gamification involves the use of structured rules and competitive conflicts to achieve specific goals, and in which a person is fully engaged with the context, losing track of time and experiencing the thrill of doing rather than achieving...
external rewards [4]. Research has shown that play and games can be linked to positive social behavior [5], the development of important skills such as self-regulation and empathy [6], and the promotion of teamwork [7].

Gamification is often used to enhance services and test gaming experiences and can be a powerful tool for influencing behavior [1]. However, some studies have examined issues related to gamification [8]. This technique is applied in settings related to learning, business, or promoting a product or service [9]. Through elements, such as competition, goal achievement, and reward, gamification can increase users’ motivation, engagement, and satisfaction in an activity [10,11].

Many researchers have suggested that gamification and game-based assessments could predict work performance as an alternative to traditional selection methods [12,13]. Fetzer et al. [13], according to the concept of “stealth assessment,” highlighted the potential of game-based assessments for predicting work performance. Invisible evaluation can accurately and efficiently detect a subject’s level of competence, can continuously extract data on their performance during the game [14], and is almost undetectable, reducing test anxiety while maintaining validity and reliability [15]. In this context, a gamified assessment environment could distract candidates from being assessed, promoting unconscious rather than desirable or socially acceptable behavior. Assessment and learning tools that may be created through the use of gamification are known as serious games [16], which are games designed for a primary purpose other than pure entertainment [17], like supporting human capital development [18]. Serious games are designed for educational, training, or informational purposes that combine the playful experience with learning specific skills and competencies [19]. Unlike traditional games, serious games were created to impart knowledge and teach players how to apply those skills in real life by creating engaging and entertaining experiences that prompt users to perform certain actions, improving the effectiveness and efficiency of those activities.

Various sectors, such as education [20], health care [21], mobility [22], marketing [23], and the environment [24], have been equipped with this innovation, simulating situations in which the player must make decisions, solve problems, or develop specific skills. Moreover, owing to the interactive nature of serious games, players can learn actively, receiving immediate feedback on their progress and the correctness of their actions [25].

If the game is user-friendly and realistic, it tends to generate positive responses from candidates [26]. Thus, one of the cornerstones of these games (as well as an aspect this paper will focus on) is the development of awareness related to one’s skills, particularly those skills called “soft skills,” which are indispensable in the world of work [27].

In relation to the connection between gamification and transversal skills, game elements, such as textures, feedback, avatars, visual effects, and voiceovers, can be effectively used to assess candidates’ soft skills [28].

Problem-solving skills, strategic thinking, and adaptation to new contexts are soft skills, also known as “behavioral skills.” They are non-technical skills (as opposed to hard skills) important for personal and professional success. Using game mechanics, gamification can make learning these skills more engaging and effective. Serious games play a key role in developing soft skills (such as problem solving, planning, and analytical thinking) in the world of work [27]. Certain elements, such as a risk-free environment, predispose users to a setting conducive to experimentation and exploration, activating stimuli, such as curiosity, learning by discovery, and perseverance [29].

Serious games are playful and interactive tools often used in education with pedagogical goals, which are designed to teach skills, notions, and strategies [30,31]. Serious games offer active involvement, immediate feedback, experiential learning, collaborative learning, and learning personalization, which enhance positive learning effects [32]. The scenarios designed for serious games have a positive relationship with student engagement and academic performance [33]. Despite their potential, serious games still pose a challenge to the educational world, and there is a need for more emphasis on less formal and traditional models that are suitable for teaching today’s generation of students [34]. Serious games are an important resource for the world of training and education, facilitating the learning of specific disciplines.

Transversal skills often overlap with digital skills, which have become increasingly necessary for performing tasks due to the growth of technology in recent years. Digital skills involve the use of information and communications technology (ICT) tools and can be classified as operational, formal, informational, or strategic skills [35].

Based on these premises, this research proposes an innovative method for addressing an issue dear to companies, the recruitment process. Recruiting an unsuitable candidate can be very costly, both in terms of money and nonproductive time. Therefore, many companies are looking for innovative ways to improve their recruitment process to attract and identify the most suitable candidates for available positions. This involves the use of cutting-edge recruiting techniques and an increased focus on assessing candidates’ skills and qualities [36].

Psychological tests that measure cognitive abilities and personality tests, which are traditional selection methods, can somewhat predict job performance [37]. As previously mentioned, the game is an attractive and valid alternative to traditional selection methods, and gamifying these tests can be a way to assess candidates with the benefits stated above.

The COVID-19 pandemic has prompted a shift in the labor market toward Industry 5.0 to the New Normal Era, a time of collaboration between advanced technology and human creativity to address business, social, and environmental challenges [38]. This has had an impact on the labor market, affecting both demand and requirements, as well as leading to the emergence of the “great resignation” phenomenon, in which individuals across countries and industries are exhibiting a heightened awareness of work conditions and opportunities [39,40].
To successfully adopt artificial intelligence (AI), organizations and individuals must embrace transformative changes in their practices, recognizing the foundational role of connected technologies in driving increased digitization in business and management worldwide. Key factors for successful AI adoption include team coordination, organizational culture, governance strategy, and AI-employee integration strategies. Furthermore, the integration of serious games and AI provides new computational tools for social research. Soft skills, such as communication, problem solving, critical thinking, and creativity, are crucial for successful AI adoption, while leadership plays a vital role in setting a vision and strategy for its implementation [41,42].

Recent studies have emphasized the importance of soft skills and talent in the New Normal Era, where work has changed significantly and new methodologies, such as smart working and remote working, are becoming increasingly prevalent. As a result, soft skills are now a top organizational priority [43].

**State of the Art**

**Industry 5.0: Impact of Emerging Technologies on Gamification and the Labor Market**

With the increasing adoption of new technologies, such as AI, virtual reality, metaverse, blockchain, augmented reality, and collaborative robots, it is crucial to understand the transition from Industry 4.0 to Industry 5.0 for creating sustainable business and management practices. Industry 4.0 was characterized by automation technologies, the Internet of Things, and smart factories, while Industry 5.0 involves the collaboration between advanced machinery and human creativity to solve business, social, and environmental issues. The adoption of transformative changes in practices is essential for organizations and individuals to understand as these connected technologies provide the foundation for increased digitization in business and management worldwide. While the use of serious games and gamification has become increasingly popular in recent years, some of these emerging technologies, particularly AI, have the potential to be disruptive for the world of work and have a significant impact on human resources, recruitment, and corporate life [38].

Team coordination and organizational culture are also crucial elements that organizations must consider when adopting AI. Effective coordination between team members is essential for successful AI implementation, and organizational culture plays a critical role in supporting this coordination. An innovation mindset is also necessary to foster a culture of experimentation, learning, and continuous improvement, which are all essential for successful AI adoption.

Governance strategy is another key area that organizations must consider when adopting AI. Effective governance ensures that AI is being used ethically, responsibly, and in compliance with legal and regulatory frameworks. AI-employee integration strategies are also important, as they help to ensure that employees feel comfortable working with AI systems and are equipped with the necessary skills to do so [41].

Pérez et al [42] suggested that the integration of serious games and AI offers new computational tools for social research. By developing and applying gamification with an eye toward both design and efficacy, organizations can increase productivity and promote better decision-making processes in the workplace. The use of gamification and serious games in the workplace can be a promising tool to enhance motivation, engagement, and decision-making processes among employees.

An evidence-based approach to gamification considers its effectiveness to be dependent on the use of scientifically proven principles and tactics that can impact outcomes [44]. Optimistic results have been found in the integration of some of the newest technologies, as AI, virtual reality, and serious games are being increasingly used in several industries, such as the healthcare industry where there was a great effort concerning new technologies and research during the COVID-19 pandemic, yielding favorable outcomes for the treatment of various health conditions as well as for medical education and training [45-47]. Moreover, Schönbohm and Zhang [48] revealed the effectiveness of serious games in facilitating strategic decision-making in the context of the COVID-19 pandemic.

In the era of the new normal, the global workforce is experiencing an unprecedented phenomenon known as the “great resignation.” This phenomenon is characterized by individuals in various countries and industries exhibiting a heightened awareness of work conditions and opportunities, which is indicative of an overall increase in people’s awareness of work. However, this phenomenon has also brought about uncertainty, challenges, and opportunities for organizations due to external factors such as the gig economy and the implementation of technology [39,40].

Recent research by Duch-Brown et al [49] investigated the relationship between market power, AI, and online labor markets. They showed that market-designing platform providers can influence labor demand and supply elasticities on online labor markets by setting specific terms and conditions on the platform. The study also revealed a significantly higher demand for AI-related labor and a significantly lower supply of AI-related labor, indicating the need for further exploration of the role of AI in online labor markets. Similarly, Deranty and Corbin [50] highlighted the potential societal impacts of AI deployment in work, particularly in the areas of technological unemployment and wealth polarization.

This presents new challenges and opportunities for soft skills [51], as well as technical skills and fusion skills [52], which are those skills and attributes required for individuals to thrive in the fusion environment of the 21st century. Therefore, staying updated with technological advancements and market-specific demands is essential for individuals employed in data science–related fields (more so than others) to enhance and maintain their employability, especially in the digital renaissance. Additionally, real-time identification of market demands plays a crucial role in balancing the adequate supply and demand of talent in this era [53].
The Future of Work Relies on Soft Skills in the New Normal Era

A new technological revolution is going on, and many organizations are eager to adopt new technologies in order to gain a competitive advantage. However, implementing new tools, like AI-based tools and technologies, is not enough to achieve success. A recent review on human resource management conducted in March 2023 revealed that organizations need to look beyond technical resources and focus on developing nontechnical resources as well [41]. The review highlights several nontechnical resources that are critical to successful AI adoption. These include human skills and competencies, such as communication, problem solving, critical thinking, and creativity, which are essential for effective collaboration with AI systems. Additionally, leadership is vital for setting a vision and strategy for AI adoption and ensuring that it aligns with the organization’s overall goals and values.

In conclusion, the review underscores the importance of developing nontechnical resources in addition to technical resources when adopting AI. By investing in human skills and competencies, leadership, team coordination, organizational culture, innovation mindset, governance strategies, and AI-employee integration strategies, organizations can fully realize the benefits of AI adoption and gain a competitive advantage in the marketplace [41].

The 2023 Pulse of the Profession report by Project Management Institute highlights that power skills, which are soft skills built upon a solid foundation of technical skills, are a top organizational priority [43]. Developing these skills can lead to delivering results that add value to both the organization and its customers.

Redefining work requires more than skill development, as skills have inherent limitations. Simply retraining workers to perform different routine tasks or adapting to new technologies to complete the same tasks does not address the underlying challenges for workers or capture the potential benefits for companies. The vision of redefined work involves shifting workers’ focus from routine tasks to identifying and resolving hidden issues and opportunities. While automation can facilitate this shift, it is not just about automating tasks or using technology. It does not involve changing the workforce composition or retraining employees for different domains, and it does not require employee suggestion boxes, 20% work time for innovation or entrepreneurship, or innovation centers. Workers must develop and use their human abilities to identify issues, solve problems, implement solutions, and iterate to recognize and resolve hidden issues and opportunities. A key characteristic of this work vision is its continuous evolution. Problem identification and solution methods are typically used to correct processes, address deviations, or eliminate inefficiencies, with the goal of incorporating feedback into more structured and defined work. Loosening the structure is only a temporary measure to advance the process [51].

The job market has undergone a significant transformation due to the pandemic, which has been expedited by the emergence of new technologies. To thrive in this new era, possessing soft skills, such as communication, problem solving, creativity, and collaboration, as well as the ability to adapt to change, is crucial [54], even if digital and technological skills are vital for enhancing a firm’s performance. Moreover, Akrim [55] stressed that while knowledge is important, soft skills cannot be overstated when it comes to securing one’s future.

To gain a competitive edge in the job market, integrating talent and knowledge management strategies is imperative [56], and to succeed in the future of work, prioritizing soft skills and talent management strategies in recruitment and training processes is essential. Identifying unique talents and skills is crucial to gaining a competitive advantage. High demand for soft skills, such as creativity, emotional intelligence, critical thinking, and problem solving, necessitates their recognition and prioritization in organizations’ development efforts [57].

The Pedagogical Side of Serious Games

Serious games serve as an effective and result-oriented medium, utilizing interactive and playful forms of engagement [30]. They represent tools designed with the goal of teaching something, such as skills, notions, and strategies, and offer the possibility for the user to identify with learning and formative situations, and thus, it can be said that they pose pedagogical goals [31]. Serious games are associated with multiple advantages from pedagogical and formative points of view, including active involvement, as the player becomes the protagonist of learning and is stimulated to explore, experiment, and reason independently [29]. Another advantage is immediate feedback, as they provide a timely and accurate response on the player’s actions, allowing learning based on error and its correction. Other advantages include experiential learning related to scientific literacy [58], as they allow for real-world situations to be experienced and simulated, providing a direct and engaging educational experience; collaborative learning, as they can be used to foster cooperative learning, stimulating participation and information sharing even in educational settings between teachers and students [59,60]; and learning personalization [61], as they can be designed to adapt to each player’s cognitive needs and abilities, sparking and maintaining player interest and motivation and enhancing positive learning effects [32]. In addition, there is a positive relationship between the scenarios designed for serious games and the expected outcomes of the learning experience [62] considering, as part of this process, student engagement and academic performance [33].

Despite the potential of these tools, they still pose a challenge to the educational world [34]. It is important to place more emphasis on less formal and traditional models that are more suitable for teaching, which can be projected to the generation of students experiencing school today. Serious games, according to this perspective, represent an important resource for the world of training and education, and are facilitators for learning specific disciplines.

Soft Skills in the Labor World

Soft skills refer to a range of personal attributes, social abilities, and communication competencies that are necessary for success in the workplace. These skills, which are distinct from technical skills, include qualities such as positive attitudes, good communication, and strong interpersonal skills. They are also
described as the “human skills” or “people skills” needed to effectively use technical knowledge and expertise in a professional setting [63].

Several studies have focused on identifying the key soft skills that are valuable in the labor market [64]. Several international studies have shown that soft skills are critical to our future success, and their importance is recognized by education providers, academics, human resource departments, and businesses [65-68].

This research aims to understand how companies and economic organizations that recruit, hire, and develop professionals at a medium to high-level view and value soft skills in their human resource development process [69].

According to the study, the most important soft skills for different groups, determined based on criteria such as significance, relevance, availability in the job market, and negative impact if lacking, are as follows: problem solving (for the group focused on achieving results), team working (for the group focused on mastering social skills), and time management (for the group focused on making progress in the workforce).

In order to examine the testability and relevance of certain soft skills, it is also useful to analyze communication and decision-making in greater depth. Additionally, problem solving may be linked to self-directed learning and communication for goal orientation, and both of these skills may be considered when discussing digital skills [70]. Therefore, it is advisable to carefully analyze these 5 skills, which were identified through a classic assessment. To assess communication [71], the communicative style (aggressive, passive, or assertive) [72-74] was adopted, and for problem solving, models such as DMAIC (Define, Measure, Analyze, Improve, and Control) [75] and Problem Solving Inventory (PSI) [76], were used. Team working and related skills developed in group environments are highly sought after by recruiters when hiring [77]. The research was inspired by the Comprehensive Assessment of Team Member Effectiveness (CATME) [78], and for decision-making, the cognitive style (a person’s preferred way of thinking and tendency to adopt certain strategies more frequently) [79] was evaluated using the General Decision-Making Styles (GDMS) scale [80]. Time management, as theorized by Lakein [81] and others, was assessed using the time management behavior scale [82]. Soft skills are increasingly recognized as critical to future success and are valued by those in education, academia, and human resource departments. They also help organizations gain new customers, increase revenue, and reduce employee turnover, which is why employers are placing an increasing emphasis on these skills [66].

Purpose of the Study and Origins
This study was conceived within the scope of the SWAG (Skills in Work-life Analysis Game) project. SWAG sets out to democratize the human resource world, through the use of intergenerational language, and it is inclusive and comprehensive of the needs, complexity, and peculiarities of today’s and tomorrow’s talents, connecting the right person to the right position. The SWAG team (involving Altomari Luca, Altomari Natalia, and Vigliaturo Daniele) was created in the context of the first cycle of “UniCaLab” (Contamination Lab of the University of Calabria) in 2019 (path concluded among the 3 winners) and achieved victory in Invitalia’s “#IoRestoAlSudHackathonTour” contest in the same year.

In this work, a new assessment framework is presented that takes into account the needs of the labor market and is based on gamification, with a focus on the concept of “recrutainment” (a term combining “recruitment” and “entertainment”), which integrates cognitive and aptitude assessment methods into the personnel selection process [83] and is not only partially conceived as a subset of gamification, but also goes beyond it [84]. In this study, a serious game that assesses the most in-demand skills in the job market and provides feedback is illustrated, proposing a set of guidelines for the development of serious games while describing the game design process. The first trial of the game was conducted and positive results were achieved, paving the way for further studies.

The serious game Among the Office Criticality (AOC) was developed by the SWAG team as a desktop version with the technical support of Artémat.

Methods

Research Methodology and Proposal of a Serious Game

The purpose of this study was to develop a model that can identify and evaluate certain work-life skills (in this case, 5 soft skills that are in high demand in the job market) and provide feedback to the user who uses the tool. Gamification is particularly suitable for this purpose because it allows for the creation of tools that can provide quick, objective, and standardized assessments [85] after an initial design and implementation effort, while also allowing for a more informal approach to the person being evaluated [12,13]. The study presents the AOC serious game, which is able to detect the 5 soft skills mentioned above. The AOC serious game uses a set of sequence analysis techniques, which is known as process mining, to overcome the limitations of traditional skills assessment methods [85]. The user interacts with the game, which tracks key performance indicators (KPIs) based on the user’s choices at each decision point and saves them in an event log at the end of the game. These events and values are processed by process mining software, which generates an output (Figure 1) [85].
The software that analyzes the event log uses an algorithm to read the data in the log and create a model. This model is then used to apply the metrics of the specific evaluation tool to produce the final report.

AOC is a serious game that is structured as a tree graph, with decisional moments represented as nodes. Analyzing the user’s choices within the game involves identifying the pattern arcs that connect the nodes. However, analyzing all user actions within the game can cause scalability issues. To address this, the event log only records data related to decisional moments and the duration of the game, rather than all interactions. In this case, sequence analysis can be helpful in solving or mitigating the problem [85]. In the AOC game, metrics are based on a data set of fixed-length numerical vectors and an overall rating formula. These vectors are added together based on the entries in the event log to create a final vector, the values of which are then used in the formula to produce a numerical result. This result, as well as the individual values, can be easily translated.
into a final report that is provided to the user in both graphical and narrative forms. The design of serious game processes can be performed in 2 steps: context analysis and iterative concept use. The first step involves using user-centered design methods, which focus on the user’s intent, situation, activity, and characteristics. The second step involves testing the tool, once it has been defined and materialized as a prototype or mock-up, in order to improve it with a focus on attractiveness, motivation, and support for activities [86]. It is important to keep in mind some fundamental principles of gamification, such as freedom of choice, benefits and relevance, personalized experience, long-term interaction, unwanted side events, and ethical and legal elements [86]. The tree graph structure containing the decisions should also be considered.

**AOC Serious Game**

AOC is a serious game that assesses 5 soft skills by immersing users in a simulated game context where they can use various tools as needed. The game context places the player directly into a company through a specific delivery at the beginning of the game. In the simulation, the player, who is starting their work experience, begins their journey at a fictitious firm with an apprenticeship contract. At the start of the simulation, the user is presented with an organizational chart, and graphic and textual descriptions of the business context. Over the first 6 months of the game, the player must interact with other characters to complete tasks, deal with unexpected events, and make decisions, choosing the means of communication and working with superiors and other team members in order to achieve the best possible result at the end of the simulation. At the end of the game, based on the choices made, the user is presented with a scenario that reflects the path they took. The contract can be converted into permanent employment, extended for an additional 6 months, or terminated.

**Serious Game Output**

In the alpha version of the game, the user is presented with a series of decisions that are related to values for various soft skills. These values are individual variables that are combined into a resultant vector, denoted as R. The R values are then input into the overall evaluation formula, which produces a numerical result, designated as x. This result corresponds to an outcome that is related to the objective of the game. As the user makes decisions, the events of the game progress according to the graph. Each event occurs consecutively and presents the user with contextual choices that are as visual and immediate as possible (eg, a choice between 3 interactive texts). The values assigned to each decision are based on data from previous studies and the literature used to construct the game [87]. The game has a maximum duration of 30 minutes, after which the available time (displayed as a countdown) reaches zero. If the simulation is not completed at this point, the final result is the R vector, which does not consider any subsequent decisional moments and includes a penalty for time management. The “first-person” perspective is intentionally designed to increase the user’s engagement and encourage active interaction.

In 2017, van Laar et al [88] conducted a study that examined 75 studies on 21st century skills and found that problem solving and communication were the most commonly mentioned skills, with 24 and 22 quotes, respectively. In this study, communication was defined as the ability to transmit information effectively, while problem solving was defined as the ability to use ICT to understand and process problems and actively use knowledge to find solutions.

The structure of the graph, the assignment of scores, and the analysis of these scores are all built using techniques from the Theory of Graphs, in line with theory and market research that emphasizes the importance of problem solving, which is tested in greater detail than the other skills. This is evident in the analysis of paths that maximize and minimize the variables associated with the skills, which has shown a range of 33 possible values for problem solving. The skills of team working, time management, and decision-making have 21 possible values and make up a total of 48% of the testing (Figure 2). The remaining 52% is composed of problem solving and communication, with this last skill requiring more attention as it is divided into 3 categories: passive communicative style, aggressive communicative style, and assertive communicative style. Variables used to identify the communicative style are also included in this percentage.
The slight differences in amplitude within the ranges can be attributed to both psychobehavioral factors and mathematical advantages. The psychobehavioral reason is that each communication style has the potential to generate unexpected events within interpersonal relationships (ordered: aggressive, passive, and assertive). The mathematical advantage is that the 3 variables can never be equal in any of the possible combinations. The outcome for the user consists of a final report that includes both graphical and narrative components (Table 1). The graphical component consists of a 2D diagram that incorporates a polygon, with each vertex representing the user’s scaled final scores in the four soft skills: team working, time management, decision-making, and problem solving. This polygon describes a surface of amplitude that is directly proportional to the x value resulting from the overall evaluation formula. The textual component includes a result relating to the final goal (3 possible scenarios) and different sections describing the user’s score in each individual skill (low, medium, and high). In the graphical component, the dominant communication style is also explicitly highlighted (aggressive, passive, or assertive).

### Table 1. Range width comparison showing “how much” a single skill is assessed.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Assessment, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem solving</td>
<td>25.19</td>
</tr>
<tr>
<td>Team working</td>
<td>16.03</td>
</tr>
<tr>
<td>Time management</td>
<td>16.03</td>
</tr>
<tr>
<td>Decision-making</td>
<td>16.03</td>
</tr>
<tr>
<td>Communication</td>
<td>26.72</td>
</tr>
</tbody>
</table>

### Ethics Approval

Ethics approval for the research project, “Innovation in Teaching Methods” was granted by the Ethics Committee of the University of Calabria (COMITATO ETICO DI ATENEO - CEA) on May 9, 2019. The project, which was conducted within the University of Calabria, aimed to assess soft skills using the tool developed by SWAG, alongside human tutors. Adhering to the approved ethical guidelines, the research team ensured participant confidentiality and privacy throughout the data analysis process.

### Results

#### Elaboration and Descriptive Statistical Analysis

To test the serious game, the researchers conducted a pilot study with 160 subjects from the University of Calabria (Italy) who were enrolled in the 2018/2019 academic year. Table 2 and Figure 3 show the results of the preliminary data screening and the normal curve of the collected data. According to the criteria of Curran, West, and Finch [89] (asymmetry greater than |1|=severe nonnormality; kurtosis greater than |3|=moderate nonnormality), the value for aggressive communication was...
close to being severely nonnormal in terms of asymmetry. This may be due to a small number of data points for individuals who had this type of communication as a result. While only 1 item showed these values, the results of the descriptive analyses can still be considered reliable.

Table 2. Normality indices related to the final scores obtained for each soft skill.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Score, mean (SD)</th>
<th>Asymmetry</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem solving</td>
<td>21.47 (3.615)</td>
<td>0.133</td>
<td>−0.109</td>
</tr>
<tr>
<td>Team working</td>
<td>13.10 (2.269)</td>
<td>−1.271</td>
<td>1.261</td>
</tr>
<tr>
<td>Time management</td>
<td>7.09 (2.251)</td>
<td>0.059</td>
<td>−0.165</td>
</tr>
<tr>
<td>Decision-making</td>
<td>12.68 (3.673)</td>
<td>−0.250</td>
<td>0.395</td>
</tr>
<tr>
<td>Assertive communication</td>
<td>10.63 (2.469)</td>
<td>−0.700</td>
<td>0.523</td>
</tr>
<tr>
<td>Passive communication</td>
<td>2.15 (2.050)</td>
<td>0.912</td>
<td>0.751</td>
</tr>
<tr>
<td>Aggressive communication</td>
<td>0.82 (1.479)</td>
<td>2.120</td>
<td>4.969</td>
</tr>
</tbody>
</table>

Figure 3. Normal distribution curves. The x-axis represents values, and the y-axis indicates probability in the normal curves.

First Try Out: Feedback Analysis

A trial of the AOC experiment was conducted, and at the end of the simulation, users were asked to complete a questionnaire about their experience. The data analysis showed a strong consistency between the user evaluation produced by AOC and the user’s self-awareness. Specifically, 88.3% (121/137) of responses to the question “Do you see yourself in the evaluation?” were at least “partly,” 70.1% (96/137) were at least “quite,” and 29.2% (40/137) were at least “yes.” It was also
useful to analyze the users’ responses to the question “How would you describe the simulation in 3 adjectives?” The most common adjective used was “interesting,” with a frequency of 16.1% (63/392) of all adjectives, followed by “fun” (36/392, 9.2%), “useful” (23/392, 5.9%), “engaging” (22/392, 5.6%), “realistic” (22/392, 5.6%), “stimulating” (17/392, 4.3%), and “innovative” (16/392, 4.1%), only considering adjectives with more than 15 repetitions (Table 3).

Table 3. Top unique repetitions of adjectives given as feedback from the assessed users to the serious game “Among the Office Criticality” after an assessment session.

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Unique repetitions (N=392), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interesting</td>
<td>63 (16.1)</td>
</tr>
<tr>
<td>Fun</td>
<td>36 (9.2)</td>
</tr>
<tr>
<td>Useful</td>
<td>23 (5.9)</td>
</tr>
<tr>
<td>Engaging</td>
<td>22 (5.6)</td>
</tr>
<tr>
<td>Realistic</td>
<td>22 (5.6)</td>
</tr>
<tr>
<td>Stimulating</td>
<td>17 (4.3)</td>
</tr>
<tr>
<td>Innovative</td>
<td>16 (4.1)</td>
</tr>
<tr>
<td>Sympathetic</td>
<td>15 (3.8)</td>
</tr>
<tr>
<td>Effective</td>
<td>10 (2.6)</td>
</tr>
<tr>
<td>Appealing</td>
<td>7 (1.8)</td>
</tr>
<tr>
<td>Constructive</td>
<td>7 (1.8)</td>
</tr>
<tr>
<td>Efficient</td>
<td>7 (1.8)</td>
</tr>
<tr>
<td>Rapid</td>
<td>6 (1.5)</td>
</tr>
</tbody>
</table>

We found it useful to group the adjectives used by users into clusters based on their focus during the simulation. These clusters were arbitrarily defined as “involvement” (cluster 1; 120/392, 30.6%), “pleasure” (cluster 2; 73/392, 18.6%), “functionality” (cluster 3; 71/392, 18.1%), “innovation” (cluster 4; 25/392, 6.4%), “truthfulness” (cluster 5; 23/392, 5.9%), “dynamism” (cluster 6; 15/392, 3.8%), “personal growth” (cluster 7; 14/392, 3.6%), “challenge” (cluster 8; 5/392, 1.3%), and “other” (43/392, 11.7%). In this study, the first, second, and third clusters accounted for 67.3% (264/392) of user responses in the survey administered after they played the serious game. We found that these clusters corresponded to other questionnaires conducted after serious games in the literature. A key factor considered, as reflected in the final feedback from users, was fidelity.

Fidelity, referring to the consistency of the game with the real world, is a crucial attribute for reality-based games as it enhances player immersion. Immersion refers to the player’s involvement in the “player’s world,” which is a crucial factor in motivating players to pursue “serious” goals [90]. However, it is important to balance immersion with entertainment elements, as previous studies have shown [27,91,92]. Fidelity can also contribute to the entertainment value of a game.

The enjoyment of playing a video game can increase motivation and lead players to find pleasure in the activities within the game, regardless of their achievements. This is especially relevant when the objectives of the game are perceived as enjoyable by the player [90].

Previous studies have found that user-perceived determinants after playing a serious game include feedback, ease of use and usability, and learning effectiveness, among others. This is consistent with the findings of this study, in which the analysis of the open-ended questionnaire taken by users at the end of the game revealed factors associated with the functionality cluster, such as ease of use and effectiveness [93]. Additionally, the enjoyment and presence-immersion factors identified in the study by Fokides et al [93] can be linked to the pleasure and involvement clusters.

According to a literature review (54 studies, 59 data sets, and 586 variables) conducted in 2019 by Baptista and Oliveira on gamification, the best predictors of acceptance for gamified technology were attitude, enjoyment, and usefulness on intention. The clusters of adjectives that were found to be most significant (involvement, pleasure, and functionality) in this study aligned with these findings, suggesting that the proposed serious game follows the trend of accepted gamification technology.

The results of the study showed that users were particularly interested in the level of engagement and interest generated by the simulation [94], as well as the entertainment value and overall emotional reactions (such as “interesting,” “engaging,” and “stimulating”). This aligns with the concept of recruitainment, which emphasizes the role of engagement in the recruitment process [95]. In addition, users were interested in the functionality of the assessment tool, as indicated by adjectives like “useful,” “effective,” and “efficient.” These results highlight the importance of designing the tool with users’ needs in mind. Overall, about half (193/392, 49.2%) of the adjectives in the involvement and pleasure clusters [86] supported the theory of user-centered design.
Discussion

Role of Gamification in Assessing Soft Skills for Talent Acquisition

This study proposes an immersive serious game model, called AOC, which aims to evaluate 5 important soft skills sought after in the job market, namely, problem solving, team working, time management, decision-making, and communication. The design of AOC was based on the literature and followed a user-centered design approach to create a user-friendly and intuitive environment. The output provided to the user holds significance as AOC was designed to generate value for both the evaluator and the evaluated subject. Therefore, it is essential to show areas of potential growth and increase the user’s self-awareness after the simulation. This study found that the proposed serious game model may be considered in line with existing literature. Nevertheless, it is worth acknowledging that this study is not without limitations. One such limitation is the relatively small sample size, which may have implications for the generalizability of the findings. To build upon and strengthen our findings, future studies with larger and more diverse samples are recommended.

Gamification is a viable solution to improve some aspects of personnel selection [95], particularly the concept of recruitainment, which focuses on the candidate’s involvement in the selection process [4,83]. Serious games are useful for both education and assessment, and are increasingly used in business and work contexts, as they help reduce selection time, avoid subjective evaluation [70], and reduce cognitive bias. They also allow for employer branding, which can positively influence the user’s opinion of the evaluator [84]. According to Diercks and Kupka [84], the use of self-assessment elements in selection can increase the success rate of achieving objectives to 93%, especially when employer branding and personnel marketing are properly carried out. The term “recruitainment,” which combines the words “recruiting” and “entertainment,” has gained popularity among Generation Y and refers to gamified, user-oriented, and simulation games that aim to improve the quality of suitable candidates through self-assessment procedures [84]. The use of gamification, combined with cognitive and aptitude assessment methods [57], is increasing in the field of recruitment and human resources in general. It is expected to triple from 2020 to 2025, with a focus on its use in personnel selection [95].

However, as noted by Kirovska et al in 2020 [95], implementing a gamification model in the recruitment process carries risks, including the need for constant monitoring. Therefore, many companies outsource it to third-party organizations to benefit from reduced hiring times and increased interaction levels, as well as to delegate risks. Attracting innovative talent is a goal for many organizations and allows them to consider characteristics and attitudes aligned with their brand’s aim. Gamification also encourages the user to do more than what is required [96], which can be a competitive advantage. In 2017, Korn et al [83] identified 42 applications (evaluating 60% of them) proposed by well-known organizations that use recruitainment, with a focus on the correlation between the profile and the hiring subject’s needs (job relatedness). They also emphasized that a low score in a single game does not necessarily mean poor user skills and that, in fact, each result could identify areas for improvement.

The human resource industry has come to fully understand and use gamification in the talent acquisition process, including from a strategic perspective. While the evaluation of technical skills is important and can be inferred from a candidate’s resume, transversal skills, which are intrinsic, innate, or developed through training and life experiences, are more difficult to objectively evaluate. This requires a significant amount of time and energy from human resource professionals [97]. In this context, gamification becomes important as it allows certain characteristics to emerge indirectly, which can be very complicated to discover otherwise [63]. Soft skills can make a significant difference in the business environment, and serious games are particularly effective at evaluating them because they can be assessed in various contexts and methods [12,13], including through self-assessment.

Conclusions and Future Developments

This research presents AOC, a serious game designed to meet the needs of the job market and place the user at the center. Evaluation techniques consistent with the literature were used to create an interdisciplinary approach for adequately evaluating the user and adding value. Initial analysis suggested that the scores obtained by subjects during the serious game can be generalized to the general population. Additionally, feedback from a postgame questionnaire indicated that the game is engaging, enjoyable, functional, innovative, and realistic, and promotes personal growth. The feedback specifically emphasizes the user journey and interest sparked by the simulation, aligning with the principles of user-centered design and recruitainment theory.

In conclusion, we propose the methodologies used in the development of the serious game AOC as a framework for designing serious games that consider the key aspects of the user journey and their relevance to theory and traditional assessment.

Future work may include analyses of the data generated by the serious game, including skills of the studied sample, skill development over time, cross-analysis with other samples, and comparison with data from other serious games evaluating different work-life skills.

Acknowledgments

We would like to thank Daniele Vigliaturo, member of the SWAG (Skills in Work-life Analysis Game) team, who actively participated in the process of design, testing, and realization of the tool, and the concept itself. We would also like to thank...
Domenico Ielasi and Angelo Mendicelli (Artémat S.r.l.) for their fundamental effort during the development and realization of the serious game Among the Office Criticality.

Conflicts of Interest
None declared.

References


44. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]


65. Dall’Amico E, Verona S. Cross-country survey on soft skills required by companies to medium/high skilled migrants. Computers in Human Behavior 2019 Nov;100:93-104. [doi: 10.1016/j.chb.2019.06.017]


70. Lin N. Assessing Classroom Participation and Performance through Gamification Systems in Foreign Language Classrooms. 2014 Presented at: Society for Information Technology & Teacher Education International Conference; March 17, 2014; Jacksonville, FL.

Original Paper

Preference of Virtual Reality Games in Psychological Pressure and Depression Treatment: Discrete Choice Experiment

Shan Jin¹*, BSc, MPhil; Zijian Tan²*, MPH; Taoran Liu³, PhD; Sze Ngai Chan⁴, MD, MPH; Jie Sheng³, MSc; Tak-hap Wong⁵, MBBS, MMed; Jian Huang⁶, BSc, MPH, PhD; Casper J P Zhang⁶, MPH, PhD; Wai-Kit Ming³, MD, PhD, MPH, MMSc

¹The Thrust of Computer Media and Art, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
²Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China
³Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China (Hong Kong)
⁴Department of Obstetrics and Gynaecology, First Affiliated Hospital of Jinan University, Guangzhou, China
⁵Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom
⁶School of Public Health, The University of Hong Kong, Hong Kong, China (Hong Kong)
*these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, PhD, MPH, MMSc
Department of Infectious Diseases and Public Health
Jockey Club College of Veterinary Medicine and Life Sciences
City University of Hong Kong
Room 1A-503, 5/F, Block 1, To Yuen Building, 31 To Yuen Street
Hong Kong
China (Hong Kong)
Phone: 852 34426956
Email: wkming2@cityu.edu.hk

Abstract

Background: Virtual reality (VR) can be used to build many different scenes aimed at reducing study-related stress. However, only few academic experiments on university students for preference testing have been performed.

Objective: This study aims to assess the preference of VR games for stress and depression treatment using a discrete choice experiment (DCE).

Methods: A total of 5 different attributes were selected based on the depression therapy parameters and attributes related to VR: (1) treatment modality; (2) therapy duration; (3) perceived remission rate; (4) probability of adverse events; and the (5) monthly cost of adding treatment to a discrete choice experiment. By comparing different attributes and levels, we could draw some conclusions about the depression therapy testing preference for university students; 1 university student was responsible for VR scene development and 1 for participant recruitment.

Results: The utility value of different attributes for “0% Probability of adverse events” was higher than others (99.22), and the utility value of VR treatment as the most popular treatment method compared with counseling and medicine treatment was 80.95. Three parameter aspects (different treatments for depression) were statistically significant (P<.001), including “0%” and “50%” of “Probability of adverse events” and “¥500” (a currency exchange rate of ¥1 [Chinese yuan]=US $0.15 is applicable) of “The monthly cost of treatment.” Most individuals preferred 12 months as the therapy duration, and the odds ratio of “12 months” was 1.095 (95% CI 0.945-1.270) when compared with the reference level (6 months). Meanwhile, the cheapest price (¥500) of depression therapy was the optimum choice for most students.

Conclusions: People placed great preference on VR technology psychological intervention methods, which indicates that VR may have a potential market in the treatment of psychological problems. However, adverse events and treatment costs need to be considered. This study can be used to guide policies that are relevant to the development of the application of VR technology in the field of psychological pressure and depression treatment.

(JMIR Serious Games 2023;11:e34586) doi:10.2196/34586

https://games.jmir.org/2023/1/e34586

JMIR Serious Games 2023 | vol. 11 | e34586 | p.671
KEYWORDS
virtual reality; discrete choice experiment; college student; depression therapy

Introduction

According to a report provided by World Health Organization [1], the incidence of depression accounts for about 4.4% of the total world population [2,3]. A detailed report [4] specifically about the prevalence of depression among Chinese individuals was recently published [5]. The China Mental Health Survey estimated that more than 95 million patients have depression in China: 83% of the patients with depression were reported to be older than 35 years old, and 65% of the patients were female [6]. Among Chinese university students, the overall prevalence of depression was 28.4% (n=185,787), and the percentage of depressed individuals was reported to be high [7].

Depression is the emotional expression of a state of ego-helplessness and ego-powerlessness to live up to certain strongly maintained narcissistic aspirations [8]. Individuals with depressive disorders usually feel sad, lonely, or irritable, and in most cases, they are in a bad mood [9]. People with depression usually experience at least five of the following 9 characteristics: (1) people feel sad and unhappy; (2) people have less interest in doing things; (3) individuals do not follow a diet, but their weight losses are obvious; (4) their weights increase, or their appetites change; (5) they suffer insomnia or drowsiness; (6) some people have mental agitation; (7) they feel that life is meaningless, or they feel guilty; (8) they have difficulty concentrating; or (9) they have suicidal thoughts [10,11].

Three categories offer reasons that individuals have this illness: (1) the influence of environmental factors (eg, abuse, criticism, or neglect in childhood); (2) the genetic factors (individuals whose close relatives have been diagnosed with depression are at an increased risk); or (3) the effects of hormones, such as serotonin and norepinephrine [12-14].

At present, the 3 main therapy methods to tackle this issue are drug treatment (antidepressants, antipsychotics), psychotherapy (psychotherapy consultation), and phototherapy (usually suitable for dealing with seasonal depression). Other alternative therapy methods are Chinese herbal medicine, exercise, meditation, self-relaxation, and others. Some antidepressants have side effects, and the cost of psychotherapy is high [15-17].

Virtual reality (VR) technology has 3I characteristics, namely, immersion, interaction, and imagination [18,19]. In the previous research on VR, as it relates to therapy for depression, a new concept of VR exposure therapy had been used in the treatment of anxiety and depression. The growing body of literature suggests that VR exposure therapy is a successful tool for relieving stress caused by psychological problems [20-22]. For example, the VR gaming experience could be used in conjunction with exercise, using games such as VirZoom, in which a VR exercise bike is compatible with most VR headsets [23,24]. The participant could wear the headset and alleviate stress and depression using the game [25].

According to the literature review, DCE is not widely applied in the field of VR technology. Research on the uses of VR technology has only covered a few areas, such as fire emergency exit testing [26], willingness-to-pay services research [27], and the testing of pedestrian behavior models [28]. By contrast, there are studies on depression and DCE. Lokkerbol et al [29] demonstrated how different components of depression treatment could impact patients’ preferences. The study conducted by van Loenen et al [30] showed that DCE could support the preference of genetic testing for the user’s willingness-to-pay and help improve depression therapy. Actually, some research suggests that the VR technology is useful in treating depression [22-25]. For example, the result from Schleider et al [31] revealed that using immersive technology, such as a VR headset, may be an efficient strategy for reducing adolescent depressive symptoms. However, to our knowledge, there is no research that investigates the participants’ preference regarding using VR technology in treating depression. There is still a big gap in the application of VR technology in DCE, and our experiments could help in this field. The purpose of this experiment is to investigate the preference for VR technology in psychological pressure relief and depression treatment for college students.

Methods

Study Sample

The study was conducted at 2 comprehensive universities in China, Jinan University in Guangzhou and the University of Electronic Science and Technology of China (UESTC) in Chengdu. We recruited our study participants from different majors at the 2 universities. Inclusion criteria included 3 parameters: (1) at least 18 years old; (2) interested in their mental state or with depressive disorders/episodes; and (3) interested in using VR technology to improve mental health problems. Those who did not agree to participate in the study and were suffering from major diseases other than depression and other mental illnesses were excluded.

Process

Figure 1 illustrates our research process. Two groups of students from different areas were enrolled in this study. Students in group 1 came from the UESTC in Chengdu, and students in group 2 came from Jinan University in Guangzhou. The students from the UESTC needed to borrow a VR device as they did not have a VR headset to support app development.
Discrete Choice Experiment

Discrete choice experiment (DCE) is a utility method in the field of economics [32]. The theory is not only applicable to economics but is also widely used in other disciplines, such as in industries, education, software, and medical care [33,34]. Medical applications include prenatal diagnostic testing and myocardial infarction diagnostic testing [35]. We conducted the questionnaire survey and VR game experience and then collected the corresponding statistical data based on the results that were collected [36,37].

VR Platform

Our devices were a Pico Neo 2 (Sichuan Shuyun Technology Co., Ltd) with a head display, earphones, handle, connecting cable, and other accessories. The VR headset also has some disadvantages: when people wear this device for a long time to play games, it may cause dizziness, nausea, eye fatigue, blurred vision, and other problems [38-40]. However, those negative influences will decrease if users just wear the headset for several minutes.

Virtual Reality Applications

Figures 2 and 3 demonstrate the VR device and application panel, respectively. It is obvious that scholars have previously performed some research experiments on static scene VR applications; thus, we selected 2 dynamic and interactive applications in this experiment: (1) VR physical and mental relaxation system and (2) VR music comfort system.

The VR physical and mental relaxation system needs participants to cooperate with the virtual professional psychologist’s guidance into scenes to help them correctly and quickly adjust to negative emotions, such as psychological anxiety and tension. First, this system could provide 16 3D realistic natural sceneries (Figures 4 and 5) so that patients can enjoy the beauty of nature without leaving home with the aim of relieving anxiety and maintaining physical and mental health for themselves. Second, based on the logic of relaxation therapy, it can provide a psychotherapist’s voice as a relaxation guide to help patients correctly relax their body and mind, and it allows them to become immersed in the atmosphere of nature to quickly eliminate negative emotions and adjust their mentality.

We designed a “VR music comfort system,” which included dozens of psychological medical music approaches. Through comfortable music activities, this system could trigger positive experiences in physiology to promote perception, memory, and arousal. This application’s key features are as follows: users can choose different scenes according to their preferences and can listen to music to relax. At the same time, users can also turn on the voice guidance function to receive guidance regarding depression and relaxation. The key step here is the user interface design because we must make sure participants could listen to the music from different VR buttons and experience the application without failure. Unity 3D and Visual Studio were used to create the user interface design, and we must make sure participants can wear the headset and click different scenes from the VR buttons correctly. Then, they can listen to music and travel to different environments in VR. At the same time, this system also integrates a positive music interaction system with the aim of promoting people’s physical and mental health; thus, music interactive scene training is carried out in the virtual world to relieve emotions [41,42].
Figure 2. Display of handle, headset, and function key in the device.

Figure 3. Application Panel for Pico Neo 2 (Sichuan Shuyun Technology Co., Ltd).
Attributes and Levels

In this paper, for the DCE attribute and the level design, we conducted a focus group with the professors in medicine at Jinan University. We came up with 5 attributes and levels. In addition, for details on the setting of the levels, we performed a literature search. There are some studies on the perceived remission rate and the incidence of adverse reactions regarding the selection of depression indicators. Different countries and experts have their own indicator considerations. Possible attributes were identified from a panel of experienced chief physicians from Jinan University. Finally, 5 different attributes were selected: (1) treatment modality, (2) therapy duration, (3) perceived remission rate, (4) probability of adverse events, and (5) the monthly cost of treatment. Based on the academic research experiments for depression therapy, we divided the attributes into several independent parameters concerning therapy duration [43,44], cost [45-48], perceived remission rate [49-55], and probability of adverse events [56-60] with the aim of building different intervals for real life. To eliminate the influence of dimensions between different attributes, the average utility values of all attribute levels were measured using the utility scaling method with zero-centered differences. Attributes and levels are shown in Table 1.
### Table 1. Attributes and levels.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Levels of attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment modality</strong></td>
<td>• kL1 Medication only&lt;br&gt;• L2 Counseling only&lt;br&gt;• L3 VRa only</td>
</tr>
<tr>
<td><strong>Therapy duration</strong></td>
<td>• L1 6 months&lt;br&gt;• L2 12 months&lt;br&gt;• L3 18 months&lt;br&gt;• L4 Lifetime</td>
</tr>
<tr>
<td><strong>Perceived remission rate</strong></td>
<td>• L1 35%&lt;br&gt;• L2 40%&lt;br&gt;• L3 45%&lt;br&gt;• L4 50%</td>
</tr>
<tr>
<td><strong>Probability of adverse events</strong></td>
<td>• L1 0%&lt;br&gt;• L2 30%&lt;br&gt;• L3 50%</td>
</tr>
<tr>
<td><strong>The monthly cost of treatment</strong></td>
<td>• L1 ¥500&lt;sup&gt;b&lt;/sup&gt;&lt;br&gt;• L2 ¥1000&lt;br&gt;• L3 ¥1500&lt;br&gt;• L4 ¥2000&lt;br&gt;• L5 ¥2500</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.<br><sup>b</sup>A currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.

### Questionnaire

This questionnaire is collected from the online questionnaire platform. It includes 4 parts: (1) basic information (5 questions: gender, education level, the impact of depression on health, level of understanding of depression, and level of understanding of VR technology); (2) the first DCE experiment; (3) VR experience evaluation feedback (5 questions: effective treatments for depression, the function of this VR app in therapy, defects of VR in psychotherapy, the best time for VR for psychotherapy, and VR scenes preferred to experience); and (4) the second DCE experiment. In the beginning, the users should complete the basic information collection step; then, they will fill in the first part of the DCE questionnaire. After finishing these sections, they will be asked to wear the VR headset and experience the apps. They will then need to finish the VR experience evaluation feedback part and the second DCE questionnaire sequence. Brief introductions for participants to acquire related knowledge at the beginning of each part, such as VR’s introduction, DCE procedure, and the meaning of the probability of adverse events, will be included. The duration of the VR applications will be around 5 minutes. Based on completing the questionnaire prior to VR application and pretesting the VR, participants were allowed 20 minutes to complete the research. An interactive design was devised; when the cursor or fingers touch the nouns (perceived remission rate, probability of adverse events), volunteers will see the introduction around the concepts and gain some useful information about them [35,36].

The first part of the DCE was conducted before the participants viewed the VR scenes, and the second part of the DCE was conducted after the participants had viewed the VR scenes. This research compared the 2 different DCE parts to understand whether VR could produce different effects on different participants. Each DCE contained 6 multiple-choice questions with 3 different options: (1) intervention mode 1, (2)
intervention mode 2, and (3) other ways. The example questions of DCE are shown in Table 2. For the DCE part, the same DCE design (ie, a random draw of the set of questionnaires from the pool with the same attributes and levels) was used for the “before-and-after” comparison. In this study, we compared the preference “before-and-after” of the participants. Our discussion about the 2 DCE parts has different questions in the “before-and-after” comparison because we subtracted various parameters from the 5 attributes in every DCE’s part to discover the user’s preference for depression therapy before and after experiencing the VR applications. Some parts of the DCE design are referred to in Ryan and Gerard’s [33] experiment, which provides better guidance on disease prevention and treatment using DCE.

Table 2. An example of discrete choice experiment question.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment modality</td>
<td>VR</td>
<td>Medication only</td>
<td></td>
</tr>
<tr>
<td>Therapy duration</td>
<td>VR</td>
<td>Medication only</td>
<td></td>
</tr>
<tr>
<td>Perceived remission rate, %</td>
<td>35</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Probability of adverse events, %</td>
<td>0</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>The monthly cost of treatment (¥)</td>
<td>500</td>
<td>2000</td>
<td></td>
</tr>
</tbody>
</table>

aVR: virtual reality.

bA currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.

Statistical Analysis

Latent Class Analysis Model

Latent class analysis was used to interpret the DCE data. The conditional logit regression model was used to quantify the correlation between the participant’s choice and the attribute levels of various test profiles for which the choices were used as the dependent variable and the attribute levels of tests as covariates. The levels for the 5 attributes were coded based on effects. The conditional logic model makes statistical inferences based on the weights of the respondents’ preferences for various attributes and levels in the questionnaire. The resulting positive and negative coefficients after the regression analysis indicated the direction of the respondent’s preference for a given attribute level. Based on the importance and magnitude of the coefficients, the marginal rate of substitution between each attribute and the relative importance of the attributes was calculated, which could measure the willingness to accept a trade-off among different options of respondents. These marginal rate of substitution values allowed for the comparison of different attributes using a common scale [61], and the odds ratios (ORs) of respondents’ preferences for different attribute levels were also reported. Sawtooth (Sawtooth Software, Inc.) was used to run the coefficients of all attributes and SEs, and t ratios to calculate the P values. For the P value of all levels of attributes, we assumed that if the P value of a level was <.05, this level was then statistically significant; when the P value of a level was <.001, this level was considered highly statistically significant.

Latent Class Model

A latent class analysis was conducted to identify correlations among explicit variables, create the fewest number of classes, and achieve local independence. According to Greene and Hensher [62], this method uses a semiparametric approach to model the correlation structure of the data and identify classes that are more homogeneous in terms of variance structure. This method of analysis can be used to sort individuals into a set of classes with a certain segmented size and scale, and different effects of each class were estimated for different attributes. In addition, this method will help us to measure the differences and similarities of preference across classes of respondents. Akaike information criterion 3 and Bayesian information criteria (AIC3 and BIC, respectively) were used as the main criteria, and after the latent class model was created, the resulting data were classified into the appropriate latent classes.

Software

All statistical analyses were performed using Sawtooth Lighthouse Studio (SSI Web version 9.11.0; Sawtooth Software, Inc.).

Ethics Approval

This study was approved by the Medical Ethics Committee of Jinan University (JNUKY-2021-004).

Results

Survey Participants

A total of 154 people were surveyed, among which 114 completed all the questionnaire contents (response rate 74%). The summary of demographic information and the suggestions and preferences of participants after using VR are shown in Table 3.
Table 3. Summary of demographic information, suggestions, and preferences of participants.

<table>
<thead>
<tr>
<th>Items</th>
<th>Total (N=114), n (%)</th>
<th>Medical students (n=71), n (%)</th>
<th>Nonmedical students (n=43), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>65 (57.0)</td>
<td>43 (60.6)</td>
<td>22 (51.2)</td>
<td>.84</td>
</tr>
<tr>
<td>Female</td>
<td>49 (42.9)</td>
<td>28 (39.4)</td>
<td>21 (48.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td>.04b</td>
</tr>
<tr>
<td>College degree</td>
<td>13 (11.4)</td>
<td>6 (8.5)</td>
<td>7 (16.3)</td>
<td></td>
</tr>
<tr>
<td>Bachelor</td>
<td>92 (80.7)</td>
<td>57 (80.3)</td>
<td>35 (81.4)</td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>8 (7.0)</td>
<td>7 (9.9)</td>
<td>1 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Doctoral</td>
<td>1 (0.9)</td>
<td>1 (1.4)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td><strong>The impact of depression on health</strong></td>
<td></td>
<td></td>
<td></td>
<td>.62</td>
</tr>
<tr>
<td>No effect</td>
<td>6 (5.3)</td>
<td>5 (7.0)</td>
<td>1 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Affects only emotions</td>
<td>22 (19.3)</td>
<td>14 (19.7)</td>
<td>8 (18.6)</td>
<td></td>
</tr>
<tr>
<td>Has a certain impact</td>
<td>29 (25.4)</td>
<td>16 (22.5)</td>
<td>13 (30.2)</td>
<td></td>
</tr>
<tr>
<td>Has a serious impact</td>
<td>57 (50.0)</td>
<td>36 (50.7)</td>
<td>21 (48.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Level of understanding of depression</strong></td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td>Not clear</td>
<td>14 (12.3)</td>
<td>8 (11.3)</td>
<td>6 (14.0)</td>
<td></td>
</tr>
<tr>
<td>Slightly heard</td>
<td>56 (49.1)</td>
<td>35 (49.3)</td>
<td>21 (48.8)</td>
<td></td>
</tr>
<tr>
<td>General understanding</td>
<td>38 (33.3)</td>
<td>24 (33.8)</td>
<td>3 (7.0)</td>
<td></td>
</tr>
<tr>
<td>Understand completely</td>
<td>6 (5.3)</td>
<td>4 (5.6)</td>
<td>4 (9.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Level of understanding of virtual reality</strong></td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Not clear</td>
<td>17 (14.9)</td>
<td>8 (11.3)</td>
<td>9 (20.9)</td>
<td></td>
</tr>
<tr>
<td>Slightly heard</td>
<td>64 (56.1)</td>
<td>43 (60.6)</td>
<td>21 (48.8)</td>
<td></td>
</tr>
<tr>
<td>General understanding</td>
<td>23 (20.2)</td>
<td>12 (16.9)</td>
<td>11 (25.6)</td>
<td></td>
</tr>
<tr>
<td>Understand completely</td>
<td>10 (8.8)</td>
<td>8 (11.3)</td>
<td>2 (4.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Effective treatments for depression</strong></td>
<td></td>
<td></td>
<td></td>
<td>.54</td>
</tr>
<tr>
<td>Psychological consultation</td>
<td>74 (64.9)</td>
<td>46 (64.8)</td>
<td>28 (65.1)</td>
<td></td>
</tr>
<tr>
<td>Medicine treatment</td>
<td>83 (72.8)</td>
<td>49 (69.0)</td>
<td>34 (79.1)</td>
<td></td>
</tr>
<tr>
<td>Movement regulation</td>
<td>57 (50.0)</td>
<td>36 (50.7)</td>
<td>21 (48.8)</td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td>67 (58.8)</td>
<td>38 (53.5)</td>
<td>29 (67.4)</td>
<td></td>
</tr>
<tr>
<td>Virtual reality treatment</td>
<td>44 (38.6)</td>
<td>21 (29.6)</td>
<td>23 (53.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Function of this virtual reality app in therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Relieve stress</td>
<td>65 (57.0)</td>
<td>36 (50.7)</td>
<td>29 (67.4)</td>
<td></td>
</tr>
<tr>
<td>Divert attention</td>
<td>71 (62.3)</td>
<td>45 (63.4)</td>
<td>26 (60.5)</td>
<td></td>
</tr>
<tr>
<td>Satisfy vanity</td>
<td>20 (17.5)</td>
<td>12 (16.9)</td>
<td>8 (18.6)</td>
<td></td>
</tr>
<tr>
<td>No function or have other function</td>
<td>14 (12.3)</td>
<td>12 (16.9)</td>
<td>2 (4.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Defects of virtual reality in psychotherapy</strong></td>
<td></td>
<td></td>
<td></td>
<td>.86</td>
</tr>
<tr>
<td>Uncomfortable wearing</td>
<td>20 (17.5)</td>
<td>12 (16.9)</td>
<td>8 (18.6)</td>
<td></td>
</tr>
<tr>
<td>Technology is immature</td>
<td>35 (30.7)</td>
<td>22 (31.0)</td>
<td>13 (30.2)</td>
<td></td>
</tr>
<tr>
<td>Expensive</td>
<td>29 (25.4)</td>
<td>17 (23.9)</td>
<td>12 (27.9)</td>
<td></td>
</tr>
<tr>
<td>Not easy to operate</td>
<td>16 (14.0)</td>
<td>8 (11.3)</td>
<td>8 (18.6)</td>
<td></td>
</tr>
<tr>
<td><strong>The best time for virtual reality for psychotherapy</strong></td>
<td></td>
<td></td>
<td></td>
<td>.47</td>
</tr>
<tr>
<td>Immediately aware of stress</td>
<td>15 (13.2)</td>
<td>10 (14.1)</td>
<td>5 (11.6)</td>
<td></td>
</tr>
<tr>
<td>When depression is diagnosed</td>
<td>27 (23.7)</td>
<td>14 (19.7)</td>
<td>13 (30.2)</td>
<td></td>
</tr>
</tbody>
</table>
For the VR experience, evaluation feedback results showed that most individuals preferred the choice of VR helmet weight, comfort, and application contents. Students also thought that VR relieves stress and diverts attention. Furthermore, students would choose the immersive scenes, memorable places from their childhood, and beautiful fantasies of the virtual world for the VR depression therapy after the VR experience.

### Attributes’ Levels and Utility Report

In the study, the utility scaling method with zero-centered differences was used to measure the average utility values of all the attribute levels and the utility gap between different treatment methods; additionally, the probability of adverse events was overt. For example, the utility of VR treatment alone was 80.95, whereas it was -42.66 and -38.30 for the attributes “Medication only” and “Psychological counseling,” respectively; for 30% probability of adverse events, the utility value was -13.81%, whereas for 50% adverse reactions, the utility value was -85.41%.

To eliminate the influence of dimensions between different attributes, the average utility values of all attribute levels were measured using the User-based Security Model with zero-centered differences. The highest utility levels of the 5 hypothesized attributes were “VR only” (of “Treatment modality”), “12 months” (of “Therapy duration”), “40%” (of “Perceived remission rate”), “0%” (of “Probability of adverse events”), and ¥500 (of “The monthly cost of treatment”; a currency exchange rate of ¥1 [Chinese yuan]=US $0.15 is applicable). The most important attribute was “Probability of adverse events,” indicating that most respondents ranked that attribute the highest (Table 4).
Table 4. The utility report of different attributes.

<table>
<thead>
<tr>
<th>Attributes and levels</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment modality</strong></td>
<td></td>
</tr>
<tr>
<td>VR&lt;sup&gt;a&lt;/sup&gt; only</td>
<td>80.95</td>
</tr>
<tr>
<td>Medication only</td>
<td>−42.66</td>
</tr>
<tr>
<td>Counseling only</td>
<td>−38.30</td>
</tr>
<tr>
<td><strong>Therapy duration</strong></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>6.62</td>
</tr>
<tr>
<td>12 months</td>
<td>23.24</td>
</tr>
<tr>
<td>18 months</td>
<td>5.72</td>
</tr>
<tr>
<td>Lifetime</td>
<td>−35.57</td>
</tr>
<tr>
<td><strong>Perceived remission rate</strong></td>
<td></td>
</tr>
<tr>
<td>35%</td>
<td>0.22</td>
</tr>
<tr>
<td>40%</td>
<td>10.81</td>
</tr>
<tr>
<td>45%</td>
<td>−5.35</td>
</tr>
<tr>
<td>50%</td>
<td>0.45</td>
</tr>
<tr>
<td>55%</td>
<td>−6.14</td>
</tr>
<tr>
<td><strong>Probability of adverse events</strong></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>99.22</td>
</tr>
<tr>
<td>30%</td>
<td>−13.81</td>
</tr>
<tr>
<td>50%</td>
<td>−85.41</td>
</tr>
<tr>
<td><strong>The monthly cost of treatment</strong></td>
<td></td>
</tr>
<tr>
<td>¥500</td>
<td>69.44</td>
</tr>
<tr>
<td>¥1000</td>
<td>16.18</td>
</tr>
<tr>
<td>¥1500</td>
<td>3.64</td>
</tr>
<tr>
<td>¥2000</td>
<td>−42.71</td>
</tr>
<tr>
<td>¥2500</td>
<td>−46.55</td>
</tr>
<tr>
<td>None</td>
<td>−336.97</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.

<sup>b</sup>A currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.

<sup>c</sup>N/A: not applicable.

**Logit Result of DCE and Attributes’ Percentage Importance**

As is presented in Figure 6, the percentage importance of the attribute “Probability of adverse events” was the highest, which shows that people reported being most concerned about the probability of adverse events. The attributes “Treatment modality” and “The monthly cost of treatment” ranked second and third, respectively.
The results of the logit analysis of all attribute levels are presented in Table 5. For treatment modality, the coefficient of the level “VR only” was positive, which indicated that the level “VR only” was positively correlated with people’s preferences and use. In addition, the coefficients of the levels “6 months,” “12 months,” and “18 months” of “Therapy duration” were positive, and “lifetime” was negative. It is clear that people showed a preference for shorter periods of therapy, but not necessarily for the shortest one. For the attribute “Therapy duration,” people were more inclined to choose the level of “12 months.” For the levels of other attributes, people preferred the lower probability of adverse events and lower treatment costs. However, for the attribute “Perceived remission rate,” people unexpectedly preferred a lower perceived remission rate rather than the “55%” option. A reasonable explanation for this phenomenon is that individuals with various degrees of histories, genders, and other backgrounds would have a different understanding of this concept (perceived remission rate). In addition, some of the participants neglected the tips around this concept in our questionnaire. We found that “VR only,” “Medication only,” and “Counseling only” for “Diagnosis methods” were highly statistically significant (P<.001 for all), “0%” and “50%” of “Probability of adverse events” were extremely statistically significant (P<.001), and “¥500” of “The monthly cost of treatment” was also statistically significant (P<.001).

Taking the level “VR only” of the “Treatment modality” attribute as the reference, the ORs of “Medication only” and “Counseling only” were 0.509 (95% CI 0.452-0.573) and 0.521 (95% CI 0.462-0.588), respectively. The OR of “12 months” of “Therapy duration” was 1.095 (95% CI 0.945-1.270) with the level “6 months” as the reference. ORs for “30%” and “50%” of the “Probability of adverse events” attribute were 0.539 and 0.364, respectively, with the level “0%” as the reference (95% CI 0.480-0.605 and 0.323-0.412, respectively). All attribute levels for “The monthly cost of treatment” were <1, which indicated that the preference weights declined with increasing medical fees.
Table 5. The result of logit analysis of preference in general (N=114).

<table>
<thead>
<tr>
<th>Attribute and levels</th>
<th>Coefficient (SE)</th>
<th>P value</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment modality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR only</td>
<td>0.44257 (0.06067)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Reference</td>
</tr>
<tr>
<td>Medication only</td>
<td>-0.23321 (0.06124)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.509 (0.452-0.573)</td>
</tr>
<tr>
<td>Counseling only</td>
<td>-0.20936 (0.06126)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.521 (0.462-0.588)</td>
</tr>
<tr>
<td>6 months</td>
<td>0.03617 (0.07616)</td>
<td>.64</td>
<td>Reference</td>
</tr>
<tr>
<td>12 months</td>
<td>0.12704 (0.07548)</td>
<td>.09</td>
<td>1.095 (0.945-1.270)</td>
</tr>
<tr>
<td>18 months</td>
<td>0.03126 (0.07691)</td>
<td>.68</td>
<td>0.995 (0.856-1.157)</td>
</tr>
<tr>
<td>Lifetime</td>
<td>-0.19447 (0.07715)</td>
<td>&lt;.05</td>
<td>0.794 (0.683-0.924)</td>
</tr>
<tr>
<td>Perceived remission rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35%</td>
<td>0.00123 (0.0894)</td>
<td>.99</td>
<td>Reference</td>
</tr>
<tr>
<td>40%</td>
<td>0.05912 (0.08968)</td>
<td>.51</td>
<td>1.060 (0.889-1.263)</td>
</tr>
<tr>
<td>45%</td>
<td>-0.02924 (0.08919)</td>
<td>.74</td>
<td>0.970 (0.814-1.155)</td>
</tr>
<tr>
<td>50%</td>
<td>0.00248 (0.08993)</td>
<td>.98</td>
<td>1.001 (0.839-1.194)</td>
</tr>
<tr>
<td>55%</td>
<td>-0.03359 (0.09085)</td>
<td>.71</td>
<td>0.966 (0.808-1.154)</td>
</tr>
<tr>
<td>Probability of adverse events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>0.54244 (0.06096)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Reference</td>
</tr>
<tr>
<td>30%</td>
<td>-0.07549 (0.05907)</td>
<td>.20</td>
<td>0.539 (0.480-0.605)</td>
</tr>
<tr>
<td>50%</td>
<td>-0.46695 (0.06215)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.364 (0.323-0.412)</td>
</tr>
<tr>
<td>The monthly cost of treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>¥500&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.37962 (0.08936)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Reference</td>
</tr>
<tr>
<td>¥1000</td>
<td>0.08844 (0.08881)</td>
<td>.32</td>
<td>0.747 (0.628-0.889)</td>
</tr>
<tr>
<td>¥1500</td>
<td>0.01993 (0.09042)</td>
<td>.83</td>
<td>0.698 (0.585-0.833)</td>
</tr>
<tr>
<td>¥2000</td>
<td>-0.2335 (0.09043)</td>
<td>.01&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.542 (0.454-0.647)</td>
</tr>
<tr>
<td>¥2500</td>
<td>-0.25448 (0.08962)</td>
<td>.005&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.530 (0.445-0.632)</td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> VR: virtual reality.<br>
<sup>b</sup> P<.001 indicates highly statistical significance.<br>
<sup>c</sup> A currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.<br>
<sup>d</sup> P<.05 indicates statistical significance.<br>
<sup>e</sup> N/A: not applicable.<br>

**Latent Class Analysis Result**

We compared these potential models, selected the model that maximized the area under the receiver–operating characteristic curve, and minimized the AIC or BIC to compensate for model complexity [63]. According to AIC, 5 classes should have been the best choice for our model. By contrast, BIC favored the 2-class option with the lowest BIC. Under such circumstances, we compared A BIC, that is, sample size–adjusted BIC, which involved sample size values. After comparison, the 2-class option had the lowest value of A BIC. Therefore, the most suitable number of latent classes in our model was 2 (Tables 6 and 7). First, all 114 respondents were divided into 2 classes with segment sizes of 77 (67.5%) and 37 (32.5%), respectively. The average maximum membership probability was around 0.87, and the percent certainty was 26.17, which is relatively low, indicating that a low level of uncertainty according to the respondents is divided into classes.
Table 6. Estimated relative preference weights for the 2 classes.

<table>
<thead>
<tr>
<th>Attribute and levels</th>
<th>Class 1 (n=77)</th>
<th>Class 2 (n=37)</th>
<th>P value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Treatment modality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR&lt;sup&gt;a&lt;/sup&gt; only</td>
<td>0.02171 (0.07483)</td>
<td>1.84767 (0.17233)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Medication only</td>
<td>−0.00327 (0.07666)</td>
<td>−1.00383 (0.14776)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Counseling only</td>
<td>−0.01844 (0.07641)</td>
<td>−0.84383 (0.14531)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td><strong>Therapy duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>0.13392 (0.09565)</td>
<td>−0.14452 (0.17659)</td>
<td>.16</td>
<td>.42</td>
</tr>
<tr>
<td>12 months</td>
<td>0.20073 (0.09506)</td>
<td>−0.13237 (0.17046)</td>
<td>.04&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.44</td>
</tr>
<tr>
<td>18 months</td>
<td>−0.06778 (0.09688)</td>
<td>0.29698 (0.17683)</td>
<td>.49</td>
<td>.09</td>
</tr>
<tr>
<td>1 Lifetime</td>
<td>−0.26687 (0.09761)</td>
<td>−0.02009 (0.17657)</td>
<td>.008&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.91</td>
</tr>
<tr>
<td><strong>Perceived remission rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35%</td>
<td>−0.16277 (0.11263)</td>
<td>0.37689 (0.19847)</td>
<td>.15</td>
<td>.06</td>
</tr>
<tr>
<td>40%</td>
<td>0.06315 (0.11437)</td>
<td>0.25085 (0.20222)</td>
<td>.58</td>
<td>.22</td>
</tr>
<tr>
<td>45%</td>
<td>−0.03595 (0.11223)</td>
<td>−0.13115 (0.20607)</td>
<td>.75</td>
<td>.53</td>
</tr>
<tr>
<td>50%</td>
<td>0.05105 (0.11274)</td>
<td>0.01580 (0.20174)</td>
<td>.65</td>
<td>.94</td>
</tr>
<tr>
<td>55%</td>
<td>0.08453 (0.11305)</td>
<td>−0.51237 (0.21644)</td>
<td>.46</td>
<td>.02&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Probability of adverse events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>0.79611 (0.07858)</td>
<td>−0.05400 (0.13866)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.70</td>
</tr>
<tr>
<td>30%</td>
<td>−0.07486 (0.07208)</td>
<td>−0.01935 (0.14090)</td>
<td>.30</td>
<td>.89</td>
</tr>
<tr>
<td>50%</td>
<td>−0.72125 (0.07926)</td>
<td>0.07336 (0.14025)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.60</td>
</tr>
<tr>
<td><strong>The monthly cost of treatment</strong>&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>¥500</td>
<td>0.29184 (0.11329)</td>
<td>0.99343 (0.20344)</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.01&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>¥1000</td>
<td>−0.01184 (0.11108)</td>
<td>0.64172 (0.20413)</td>
<td>.91</td>
<td>.002&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>¥1500</td>
<td>−0.01581 (0.11449)</td>
<td>0.02437 (0.20354)</td>
<td>.89</td>
<td>.90</td>
</tr>
<tr>
<td>¥2000</td>
<td>−0.18010 (0.11332)</td>
<td>−0.39729 (0.21176)</td>
<td>.12</td>
<td>.06</td>
</tr>
<tr>
<td>¥2500</td>
<td>−0.08408 (0.11113)</td>
<td>−1.26223 (0.23087)</td>
<td>.45</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.
<sup>b</sup>P<.001 indicates highly statistical significance.
<sup>c</sup>P<.05 indicates statistical significance.
<sup>d</sup>A currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.
Table 7. Odds ratios and CIs of attribute levels in the 2 classes.\(^a\)

<table>
<thead>
<tr>
<th>Attribute and levels</th>
<th>Class 1 (n=77), odds ratio (95%CI)</th>
<th>Class 2 (n=37), odds ratio (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment modality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR(^b) only</td>
<td>Reference 0.975 (0.842-1.129)</td>
<td>Reference 0.058 (0.041-0.081)</td>
</tr>
<tr>
<td>Medication only</td>
<td>0.961 (0.827-1.116)</td>
<td>0.068 (0.051-0.090)</td>
</tr>
<tr>
<td>Counseling only</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Therapy duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>Reference 1.069 (0.887-1.288)</td>
<td>Reference 1.012 (0.725-1.414)</td>
</tr>
<tr>
<td>12 months</td>
<td>0.817 (0.676-0.988)</td>
<td>1.555 (1.100-2.199)</td>
</tr>
<tr>
<td>18 months</td>
<td>0.670 (0.553-0.811)</td>
<td>1.133 (0.801-1.601)</td>
</tr>
<tr>
<td>Lifetime</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived remission rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35%</td>
<td>Reference 1.253 (1.002-1.568)</td>
<td>Reference 0.882 (0.593-1.310)</td>
</tr>
<tr>
<td>40%</td>
<td>1.135 (0.911-1.415)</td>
<td>0.602 (0.402-0.901)</td>
</tr>
<tr>
<td>45%</td>
<td>1.238 (0.993-1.545)</td>
<td>0.697 (0.469-1.035)</td>
</tr>
<tr>
<td>50%</td>
<td>1.281 (1.026-1.598)</td>
<td>0.411 (0.269-0.628)</td>
</tr>
<tr>
<td>55%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Probability of adverse events</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>Reference 0.419 (0.363-0.482)</td>
<td>Reference 1.035 (0.785-1.365)</td>
</tr>
<tr>
<td>30%</td>
<td>0.219 (0.188-0.256)</td>
<td>1.136 (0.863-1.495)</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The monthly cost of treatment(^c)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>¥500</td>
<td>Reference 0.738 (0.594-0.918)</td>
<td>Reference 0.703 (0.472-1.050)</td>
</tr>
<tr>
<td>¥1000</td>
<td>0.735 (0.587-0.920)</td>
<td>0.379 (0.255-0.565)</td>
</tr>
<tr>
<td>¥1500</td>
<td>0.624 (0.500-0.779)</td>
<td>0.249 (0.164-0.377)</td>
</tr>
<tr>
<td>¥2000</td>
<td>0.687 (0.552-0.854)</td>
<td>0.105 (0.067-0.165)</td>
</tr>
<tr>
<td>¥2500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The table presents general results of the multinomial logit model. Data on preferences of respondents from the 2 classes for psychotherapy intervention attributes are reported (N=114).

\(^b\)VR: virtual reality.

\(^c\)A currency exchange rate of ¥1 (Chinese yuan)=US $0.15 is applicable.

For class 1, the attribute “Probability of adverse events” was the most important factor for respondents with a percentage importance value of 55.29%, followed by “The monthly cost of treatment” and “Therapy duration,” with percentage importance values of 17.20% and 17.04%, respectively (Figure 7). Meanwhile, the span of percentage weights of “Probability of adverse events” was obvious (Figure 8), ranging from −0.72 to 0.79. The span of the attribute “The monthly cost of treatment” ranged from −2.33 to 0.29. The preference weight of “The monthly cost of treatment” decreased with increasing expenses. In addition, the OR of “50%” of the attribute “Probability of adverse events” was 0.219 (95% CI 0.188-0.256), indicating that the majority of patients preferred to pay more attention to the safety of treatment modality with respect to all the levels of the attribute “The monthly cost of treatment” when compared with the reference level ¥500 of <1. At the same time, the OR (Table 7) of the levels of the attribute “Therapy duration” decreased with treatment time, indicating that peoples’ preference for this attribute increased as treatment time decreased.

For class 2, the attributes “Treatment modality” and “The monthly cost of treatment” were relatively important; for respondents in this class, the percentage importance values were 43.43% and 34.36%, respectively (Figure 7). For the attribute “Perceived remission rate,” the percentage importance value was 13.54% (Figure 7). The span of the percentage weights of these 2 attributes is presented in Figure 3. The percentage weights for the attributes “Treatment modality” and “The monthly cost of treatment” ranged from 0 to −1.003 and −2.33, respectively (Figure 9). In addition, the OR of “Medication only” of the attribute “Treatment modality” was 0.058 (95% CI 0.041-0.081), which was <1, indicating that the VR method appeared to be the best of the 3 methods. The OR of the levels of “Therapy duration” were all >1, and it was...
different from class 1. Meanwhile, the OR continued to decrease with the increase in the monthly cost of treatment, which was the same condition as that found in the previous class.

For these 2 latent classes, the attribute “Probability of adverse events” was the most preferred factor in class 1, whereas “Treatment modality” was the most preferred factor in class 2. There were 74 and 34 participants in class 1 and class 2, respectively. Class 1 has a similar number of male (39/74, 52.7%) and female participants (35/74, 47.3%), whereas the number of male participants (21/34, 61.8%) is nearly 2 times than that of female participants (13/34, 38.2%) in class 2. In addition, nearly two-thirds of participants in class 1 were medical students (48/74, 64.9%), with nonmedical students comprising only a minority (26/74, 35.1%). By contrast, the numbers of medical and nonmedical students in class 2 were almost even (18/34, 52.9%, and 16/34, 47.1%, respectively).

**Figure 7.** Percentage importance of attributes in the latent class condition.

**Figure 8.** Latent class percentage weights in class 1. VR: virtual reality.
Discussion

Principal Findings

Through the investigation of 5 different attributes and several levels in this survey, our preliminary results indicate that the treatment modality and the probability of adverse events will interfere with people’s choice of therapy. Although a host of people are willing to choose VR as the treatment method instead of others, the treatment modality is not the element that most concerns them. However, the influential weight of the treatment modality is second only to the probability of adverse events, especially for college students who have used VR devices. Although the VR headset did not cause severe adverse reactions immediately, such as vomiting and diarrhea, the disadvantages of overweight devices and dizziness during use should not be ignored.

People who have previously used a VR headset and have had uncomfortable reactions after experiencing VR may have different preferences than those reported in this survey. After we finish the DCE experiment for university students, we hope to carry out a clinical VR experiment for the treatment of depression. Although VR has many advantages in the treatment of mental diseases, some problems may occur when VR is actually applied in the treatment of severe psychological depression, such as a lack of willingness to wear VR glasses. We are willing to compare the treatment of VR in clinical patients to better improve the use of VR in stress counseling and depression treatment.

Latent class analysis results showed that most participants in class 2 would consider “Treatment modality” as the most important attribute. As the number of male participants is almost 2 times more than that of female participants in class 2, this result may be due to a more receptive attitude toward VR technology and VR therapy among male participants, compared with female participants. Male students are also more likely to play computer games and VR games in daily life, whereas women might pay more attention to the details of the equipment, such as equipment weight, dizziness, and air permeability. In addition, our results revealed that the majority of participants in class 1 would consider the “Probability of adverse events” as the most important attribute. This result is possibly due to the prominent number of medical students in class 1. Because of the specialty of their major, it is common that participants from clinical medicine and related majors would understand more about the side effects of antidepressants and be more aware of the complications of using VR technology. Thus, it is not surprising that most participants in class 2 chose “Probability of adverse events” as the most important attribute.

The results of the DCE experiment in this study are helpful for doctors and researchers to conduct further analysis, so they can devise better clinical strategies for the treatment of depression. We separated different treatment methods and compared the effects of individual choices, such as just focusing on medical therapy for depression and analyzing the influence from different perspectives (time and cost). At the same time, our experimental results could promote VR technology for the treatment of depression and provide reference and guidance for individuals and hospitals that are willing to use VR treatment.

Our study’s results showed that individuals prefer VR technology as a psychological intervention method for treating psychological disorders (such as depression) and easing their anxiety. There are several studies that might support our findings [30,31,64,65]. First, an article suggested that a virtual and...
peaceful environment with relaxing music could make the participants feel like they are in a realistic soothing environment [64]. Second, the research done by Ioannou et al [65] stated that VR could effectively reduce these symptoms in different contexts and with regard to different diseases, including cancer. Another paper suggested that VR has good clinical potential to reduce the severity of depression [30]. A study conducted by Schleider et al [31] revealed that immersive technology, such as VR headsets, may be an efficient strategy for reducing adolescent depressive symptoms. According to the aforementioned studies, the comfortable feeling that VR technology can provide and its effectiveness in treating depression may explain the participants’ preference for using VR technology in treating psychological disorders.

Indeed, most other VR depression experiments are focused on the methodology and VR content making more than the university students’ depression therapy. By contrast, participants’ preferences in our results, such as “Probability of adverse events” and “Treatment modality” from the DCE, could help doctors adopt some new strategies with VR headsets for depression therapy. Besides, in the university environment, the mental health treatment center and psychological counseling room in the school could use VR applications to prevent depression and anxiety for students. Another similar study used VR to enhance employees’ mental health and work performance [66].

**Limitation**

There are some limitations to this study. First, the new high technocracies that involve VR may influence university students’ preferences. In addition, participants with knowledge of drugs (antidepressants), such as selective serotonin reuptake inhibitors and serotonin and norepinephrine reuptake inhibitors, may cause bias in the preferred choice in the questionnaire [67]. All of these factors will cause the experimental process to become more complex. We will continue to perform a variety of mixed treatment experiments in the future to further accurately analyze the effects of different combinations of treatment methods for depression. At the same time, we will also focus on more parameters, such as gender, major, and age, in depression therapy. For example, medical students will have some bias in depression treatment compared with other students.

**Conclusion**

People placed significant preference on VR technology’s psychological intervention methods, which indicates that VR may have a potential market for treating psychological problems, especially for university students. However, most participants were concerned about the side effects and treatment modality of the VR technology, and these 2 attributes need to be considered when designing treatment plans for individuals. This study can guide policies relevant to the development of the application of VR technology in the field of psychological pressure and depression treatment. At the same time, doctors can consider adopting multiple methods, such as VR plus counseling, in clinical applications.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

AIC3: Akaike information criterion 3
BIC: Bayesian information criteria
DCE: discrete choice experiment
OR: odds ratio
UESTC: University of Electronic Science and Technology of China
VR: virtual reality

©Shan Jin, Zijian Tan, Taoran Liu, Sze Ngai Chan, Jie Sheng, Tak-hap Wong, Jian Huang, Casper J P Zhang, Wai-Kit Ming. Originally published in JMIR Serious Games (https://games.jmir.org), 16.01.2023. This is an open-access article distributed under

https://games.jmir.org/2023/1/e34586
JMIR Serious Games 2023 | vol. 11 | e34586 | p.690
(page number not for citation purposes)
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.