JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 8 (2020), Issue 3 ISSN 2291-9279 Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Viewpoint

Original Paper

Developing a Theory-Driven Serious Game to Promote Prescription Opioid Safety Among Adolescents: Mixed Methods Study

Olufunmilola Abraham[1], BPharm, MSc, PhD; Tanvee Thakur[1], BPharm, MSc; Randall Brown[2], FASAM, PhD, MD

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

[2]Department of Family Medicine and Community Health, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Olufunmilola Abraham, BPharm, MSc, PhD
Social and Administrative Sciences Division
School of Pharmacy
University of Wisconsin-Madison
Room 2515 Rennebohm Hall
777 Highland Avenue
Madison, WI, 53705
United States
Phone: 1 608 263 4498
Fax: 1 608 262 5262
Email: olufunmilola.abraham@wisc.edu

Abstract

Background: Adolescents in North America are severely affected by the opioid crisis, yet there are limited educational resources for educating teens about prescription opioid safety and misuse. Empirical literature lacks evidence regarding teen education about prescription opioid safety through serious games and lacks conceptual models and frameworks to guide the process of game development for this purpose.

Objective: This study aims to conceptualize and design a serious game prototype to teach teens about prescription opioid safety and propose a conceptual framework for developing a serious game to educate youth about safe and responsible use of prescription opioids.

Methods: The initial steps of the project comprised of the formulation of an integrated conceptual framework that included factors from health behavior models and game development models. This was followed by the formal process of serious game development, which resulted in a game prototype. The assessment of the game prototype was done through group discussions, individual interviews, and questionnaires with adolescents following gameplay. Field notes were used to keep track of the responses from the group discussions. Content and thematic analyses were used to analyze field notes and responses to the open-ended questionnaire, which were then used to refine the game prototype.

Results: A total of 10 playtests with over 319 adolescents and emerging young adults (AYAs) in community settings such as middle schools, high schools, and colleges were conducted by the project team between March and June 2019. The AYAs provided feedback on the initial game prototype using questionnaires administered through Qualtrics or in-person on paper. Preliminary feedback suggested that the teens found the game objectives, outcomes, and design appealing. Overall, the game was perceived as realistic, and learning outcomes seemed achievable. Suggestions for improvement included the need for additional direction on gameplay, clearer instructions, concise dialog, and reduced technical problems in the gameplay.

Conclusions: We propose a conceptual framework for developing a serious game prototype to educate youth about prescription opioid safety. The project used a theory-driven conceptual framework for the development of a serious game targeting the prevention of adolescent opioid misuse and garnered preliminary feedback on the game to improve the quality of gameplay and the prototype. Feedback through informal assessments in community settings suggests that the youth and their families are interested in a game-based approach to learn about prescription opioid safety in homes and schools. The next steps include modifications to the game prototype based on feedback from the community, integration of learning analytics to track the in-game behaviors of players, and formal testing of the final prototype.

(JMIR Serious Games 2020;8(3):e18207) doi:10.2196/18207

KEYWORDS
opioid; medication adherence; adolescents; youth; video games; games; mobile phone; educational technology

Introduction

Background

The misuse of prescription opioids and the associated negative health outcomes is a leading public health issue among teens in the United States [1]. Over the past 20 years, there has been an increase in the prescription and misuse of opioids among teens, as well as an increase in overdoses and overdose fatalities [1-3]. There is a significant risk for teens to experience an adverse drug event (such as sedation, physical dependence, or overdose) from either intentional or unintentional use of an opioid prescription [2]. In 2016, 11.8 million individuals aged 12 years or above living in the United States misused opioid medications [2]. Approximately 3.6% of adolescents aged between 12 and 17 years were reported to misuse opioids in 2016 [2]. Many teens and their family caregivers do not have a good understanding of the prescribed medications, particularly opioids [4]. Studies have shown that many individuals are not well educated about the side effects of opioid medications and the associated risks of improper storage. This lack of knowledge leads to adverse patient-centered events [4-9]. Hence, it is essential that teens receive developmentally appropriate education on opioid medication safety. Adolescent-focused education on the dangers of misusing or abusing prescription drugs, proper storage, and disposal of medications is increasingly recommended and is key in addressing inappropriate opioid use in this population [1].

Currently, there are few regulations or interventions to guide the education about opioids, particularly for teens [9]. Paper pamphlets are one of the few tools that exist; they tend to be theoretical and didactic in format and are neither patient-centered nor engaging. Educational materials that are intended to provide medication information to the youth and their families are at an 11th-grade reading level, which makes them difficult for all teens to comprehend [10-15]. Suboptimal educational materials and lack of awareness of the risks of opioid diversion and addiction in teens likely contribute significantly to the increase in nonmedical opioid use, misuse, addiction, overdoses, and deaths in this population [16-18].

Understanding and acting upon the existing medication information and instructions pose significant obstacles for almost half of the Americans [19,20]. Although a variety of opioid educational programs are available for adults, a digital tool designed to address the opioid medication safety knowledge gap in teens does not currently exist. Technology-based *serious games* are a novel method of delivering interactive health behavior education through skill-building exercises [21,22]. Serious games are digital tools that offer engagement activities through a responsive narrative to educate the participants through role-play and practicing skills. Unlike traditional video games, serious games act to convey meaningful information through interactive environments similar to real-life situations [23,24]. Serious games offer methods for increasing knowledge, delivering persuasive messages, changing behaviors, and influencing health outcomes among adolescents [1,25].

Although games for health offer promising positive patient outcomes [26-28], there is a dearth of research on the use of serious games to provide developmentally appropriate prescription opioid education for teens. A recent systematic review of serious games that incorporated medication usage identified only 12 serious games (targeting adolescents, young adults, and adults), with a majority aiming to increase adherence to specific medications or for a specific medical condition (such as cancer, diabetes, and asthma). Findings from the systematic review suggest an opportunity to use serious games to improve opioid medication safety and focus on preventing adverse drug events [29].

Behavioral theories, conceptual models for game development, and playtesting are known to improve the effectiveness of serious games [30,31]. There is limited evidence available about the use of specific theoretical frameworks to guide the game development process for medication safety–focused games for youth [29,32]. A few serious games about medication adherence and education have incorporated varying degrees of theoretical frameworks in their game design and assessment, but these games focus more on medication adherence than education about medications and effects [21,30-33]. Serious games that adopted theory in their design used small sample sizes for testing efficacy and effectiveness; hence, positive outcomes were demonstrated by only a few of these studies [30,34]. These studies applied either a gamification theory or the social behavioral theory to develop games incorporating medication usage as learning outcomes. To our knowledge, an integrated conceptual framework that includes concepts from both gamification and health behavior theories has not been developed and used to create a developmentally appropriate serious game for prescription opioid safety education for teens.

Objectives

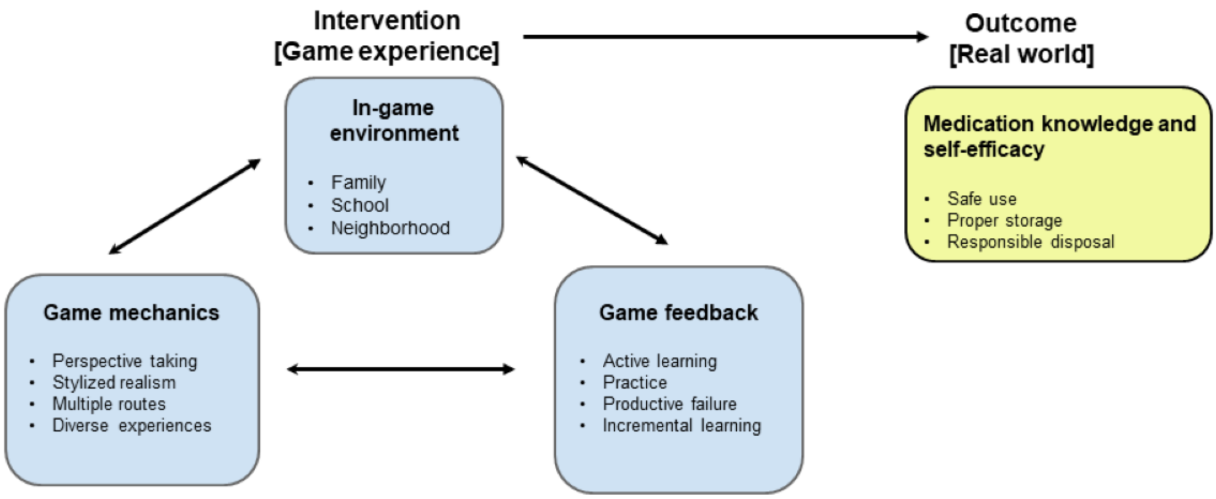
The *MEDSMAR$_x$T: Adventures in PharmaCity* intervention was designed as a serious game to provide interactive education about prescription opioid risks and safety for adolescents. The goal of this game is to improve the knowledge, awareness, and self-efficacy teens have on the safe, appropriate, and responsible use of prescription opioid medications. Given the need for developing serious games to educate teens about prescription opioid safety and the lack of theory-driven conceptual models available in empirical literature to guide the process of game development, this project aimed to conceptualize and design a serious game prototype to educate adolescents about prescription opioid medication safety and propose a conceptual model for developing and assessing the use of the serious game in community-based settings. Community-based settings, in this context, include educational settings, such as schools, universities, and youth clubs, as well as health care settings in the community, such as hospitals, clinics, and pharmacies.

Methods

Conceptual Framework

A serious game behavior change framework for improving medication safety in community settings was developed (Figure 1) as the theoretical approach to guide this project. The proposed framework was based on behavioral and psychological interventions that incorporate a broad range of technologies, such as serious games, to positively change behaviors and cognitions related to health, mental health, and wellness [35]. The game intervention comprises 3 domains: the in-game environment, game mechanics, and game feedback, which are cyclically related to each other. The conceptual framework is derived from social behavioral and game development theories. Thus, the game experience is expected to lead to real-world outcomes (Figure 1).

Figure 1. Serious game framework for improving medication safety in community settings.



The framework was informed by the principles of behavioral theories such as social cognitive theory and social learning theory. The principles from these models guided the selection of relevant constructs for the game experience and measurable learning outcomes such as medication self-efficacy [36,37]. The social cognitive theory posits that the combination of personal and environmental factors is a key influencer of behavior [38]. Each of these influencers was examined when creating this model, and game mechanisms that support them were identified. These mechanics were primarily informed by the learning principles from the foundational text by Gee, *What Video Games Have to Teach Us About Learning and Literacy*, and the literature on good instructional design practices [39-41]. The social learning theory proposes that individuals can acquire new behaviors by learning and imitating the behaviors of people with significant influences in their lives. In the case of adolescents, these include their parents, siblings, and friends. Thus, the in-game environment and the situations involved in the gameplay (such as events at home and school) were based on the principles of social learning and social cognitive theories [38]. The bidirectional interaction between personal and environmental factors and influential people who function as models for adolescents were considered significant locations and scenarios for the in-game environments (home or family, school, and neighborhood). The primary outcomes of *increased medication self-efficacy* in the conceptual model were informed by the social cognitive theory [38].

The behavioral theories were supported by the components of the behavioral intervention technology model, which integrates the principles of behavioral science, technology design, and engineering [42]. This theoretical framework defined both the conceptual and technological architecture of the game-based intervention by identifying the key concepts related to behavior change strategies such as education, design elements of technology, and characteristics of the interventions such as the needs, capabilities, and preferences of a user. Conceptually, the framework emphasizes the promotion or reduction of specific behaviors, the technology design characteristics, and the behavior change strategies.

Consistent with the theoretical dimension of the framework, the content of the serious game intervention will increase knowledge and teach adolescents essential skills to improve their self-efficacy in safe medication use, proper medication storage, and proper medication disposal, which are the expected outcomes of this intervention. The different terms used in the framework are given below.

The terms related to the game mechanics are as follows:

- Perspective taking: During gameplay, learners can take on identities to experience novel situations and roles, picture themselves in those settings, and build empathy toward the situation [39].
- Stylized realism: In game terminology, stylized realism involves using rich, authentically designed scenarios to support long-term learning [43].
- Multiple routes: Giving learners multiple options to explore empowers them and provides opportunities for replaying content. Showing different consequences of actions is powerful feedback [40,41].
- Diverse experiences: Simulations and games can provide a wide variety of experiences that are real enough to matter but still safe to explore [44].

The terms related to the game feedback are as follows:

- Active learning: In contrast to passive learning, active learning is a broad category that includes any instructional technique that engages the student in thoughtful, meaningful learning [45].
- Practice: Repeating attempts to recall and/or apply knowledge promotes long-term recall and transfer [46].
- Productive failure: This design strategy leverages short-term failure for long-term learning gains as learners come up with multiple attempts at solutions [47].
- Incremental learning: Learners have repeated attempts to produce a sense of challenge and develop learner persistence as well as repeated exposure to the learning material and relevant cues [43].

The framework suggests that the game-based intervention, which is a serious game experience, will have a direct effect on the learner's knowledge and behaviors such as safe medication usage, proper medication storage and disposal, and educating peers or family members about opioid safety. Within the serious game experience, environmental factors where adolescents may interact are deliberately represented through game design and mechanics, such as support networks (peers), family (home environment), neighbors, and the school or city bus.

Game Development Process

Creating a Multidisciplinary Team

The first step was to assemble an experienced project team with diverse and complementary expertise and experience in medication safety, adolescent health, opioid use, and game development. The project team comprised pharmacists, physicians, social behavioral health experts, student pharmacists, teens, game designers and developers, and researchers with expertise in improving medication use and health outcomes. The game developers were chosen based on their prior experience in developing serious games.

Knowing Your Audience

We assessed the baseline knowledge of the target population about prescription opioid medication misuse and safety as well as existing educational platforms preferred for learning about prescription opioid safety. This was conducted through formal and informal discussions with teens using focus groups, individual interviews, and questionnaires in the community. The target audience for this preliminary testing included middle school, high school, and college students.

Creating Learning Outcomes of the Game

The project team met weekly for 1 year (May 2018 through May 2019) to develop the first version of the serious game (MEDSMAR$_x$T: Adventures in PharmaCity) prototype. The team meetings focused on identifying, discussing, and finalizing feasible and attainable learning goals for the game. The learning outcomes of the game were drawn from 8 principles of opioid safety agreed upon by the project team and informed by various opioid medication safety resources [48-50]. Players are expected to learn prescription opioid safety principles through gameplay and how to apply them in real-life situations. This includes safe usage, proper storage, and responsible disposal.

Applying the Conceptual Model

The MEDSMAR$_x$T: Adventures in PharmaCity game aimed to improve knowledge, attitudes, and behaviors, and self-efficacy of adolescents regarding the safe and responsible use of prescription opioids. To meet this overarching goal, the project team created a conceptual framework comprising both gaming and health behavior theories. The conceptual framework entitled *a serious game behavior change framework for improving medication safety in community settings* (Figure 1) provided the theoretical underpinnings that guided this project. This framework was used to conceptualize learning objectives, gameplay outcomes, and game scenarios.

Creating a Game Playbook

The project team exchanged and shared ideas and knowledge about the behavioral and clinical health aspects and implications for game design and prototype development. A game playbook was created to serve as a guide for game design and development. It incorporated ideas from members of the multidisciplinary project team and was accepted by all on the team. This game playbook was critical to the team at different stages of prototype development to create and refine the content and to serve as source material for the preliminary informal testing of the game. The playbook also served as a reference resource for all team members for protocols and procedures.

Using an Iterative Feedback Process for Game Development

During weekly meetings, the team reviewed the progress of the game and provided feedback. Feedback was provided on domains (game environment and mechanics) of the conceptual model to ascertain that learning objectives were met, if the game was clinically relevant, and on the physical characteristics of the game (characters, aesthetics, and gameplay). The game was revised for 1 year, based on this continuous feedback and using an iterative process, until a playable game prototype was ready for gameplay and testing in community settings.

Game Synopsis

The project team agreed on the use of an animated, animal-based character design for the MEDSMAR$_x$T: Adventures in PharmaCity prototype game. Anthropomorphized characters were used, which may interest a diverse audience of gameplayers irrespective of gender, age, race, or ethnicity to easily identify with the game characters. The initial game prototype included animal characters, such as an anthropomorphized sheep as the main character (Figure 2). The anthropomorphic style of the game is reminiscent of currently popular games that target youth. Within the game, anthropomorphic characters interact within different environments and communicate between player and nonplayer characters using an in-game cell phone dialog mechanism (Figure 3). The use of such visually attractive images is supported in the literature as a means of connecting players with game synopsis, characters, and gameplay [51]. The project team emphasized using *animated stylized realism* that supported scenarios authentic to the real-world experiences of the youth. Rich, authentically designed scenarios support learning [52] and long-term retention of information and knowledge [53].

The game prototype design includes several locations that are reflective of community settings where youth might interact (ie, at home, in school, or on a bus). These situations were selected to highlight realistic settings where teens may engage in medication usage discussions and face ethical dilemmas related to the safe and responsible use of medicines. In each setting, players are asked to think from the lens of someone going through a situation involving opioid safety practices, thus implementing *perspective taking* in gameplay. Each situation highlights different challenges and requires decision making and critical thinking. *Active learning* is used here such that players in the game are faced with the positive and negative consequences of their choices as they explore options and develop strategies for navigating medication usage, storage, disposal, and related social pressures. Players could proceed to the next step only if they could successfully complete the previous level. Thus, the *practice* of opioid safety behaviors was implemented in the game in realistic scenarios such that players could also use them in the real world if needed. In the home scenarios, the player must discover how to safely store and dispose of prescribed opioids to avoid negative outcomes such as friends getting sick from finding and misusing these medications (Figure 4). In the bus scenario, the player is faced with a dilemma about sharing medications, accidental overdose, and naloxone use (Figure 5). At school (Figure 6), the player encounters another choice regarding prescription opioid misuse. All these different situations in the game provide *diverse experiences* for the gameplay.

In these scenes, the player must choose how to respond, with options such as texting friends and family for advice. This provided an option of *multiple routes* for players where they could use one or more routes to decide and complete a level in the game. Players have an opportunity to *turn back the clock* and repeat scenes until they navigate them successfully (Figure 7), thereby learning from *productive failures* as well as successes. Repeated attempts also produce a sense of challenge and develop player persistence as well as repeated exposure to the learning material and relevant cues [41]. This *incremental learning* prompts learners to think back to the simulated situations when faced with analogous ones in the real world. This level of control and freedom is unique to interactive media tools such as serious games and promotes engagement and learning [39,41,43].

Figure 2. Anthropomorphized characters in the MEDSMAR$_x$T game.

Figure 3. In-game cellphone dialogue.

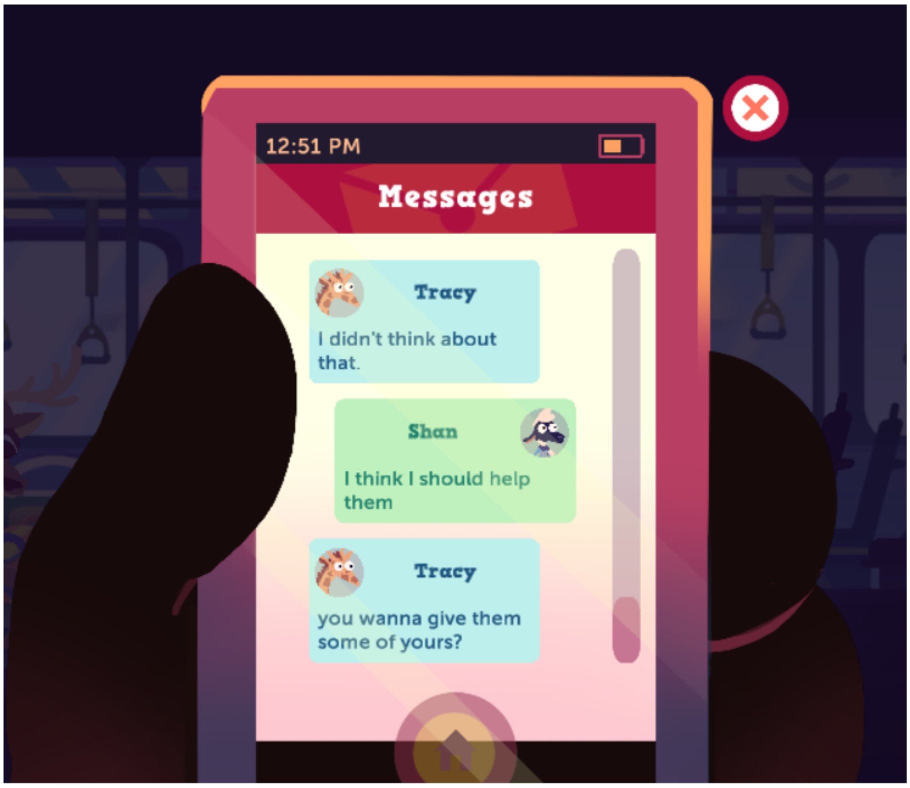


Figure 4. Home scenario in the game.

Figure 5. Bus scenario in the game.

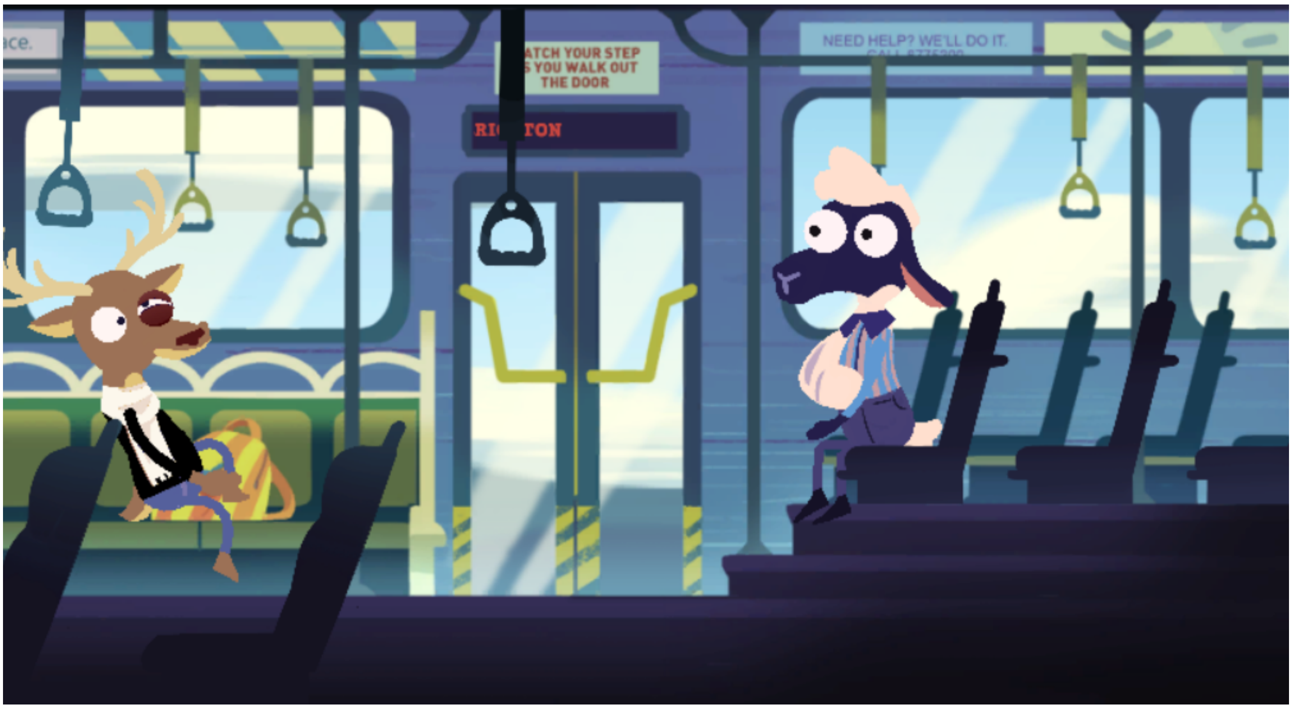


Figure 6. School classroom scenario in the game.

Figure 7. Repeating gameplay feature.

Results

Assessment and Evaluation of the Game Prototype

A total of 10 informal playtests with 319 adolescents and emerging young adults (AYAs) from middle schools, high schools, and colleges were conducted by the project team between March and June 2019. The participant sample was obtained by approaching school districts, universities, and youth clubs across the state of Wisconsin. Gameplay was conducted as a part of the activities in classroom sessions if allowed by the schools, universities, and youth clubs. The sites for preliminary testing and their characteristics are described in Table 1. The goal of the playtests was to obtain feedback to improve the quality of the game prototype related to game dynamics, design, and adaptability. Participants played the game individually and in groups. Feedback was obtained using individual questionnaires and through group discussions.

Table 1. Playtest sites and characteristics reported.

Site	Age (years)	Participants	Number of playtests
Middle school	11-13	Overall, 41.4% (132/319) students; 60.6% (80/132) male; race/ethnicity: 59.8% (79/132) white, 13.6% (18/132) Latino, 8.3% (11/132) African American, and 17.4% (23/132) Asian	6
High school	15-18	Overall, 4.4% (14/319) students; 7% (1/14) female; race/ethnicity: 79% (11/14) white and 7% (1/14) Asian	1
High school	15-17	Overall, 24.5% (78/319) students	3
Public library	14-19	Overall, 5.3% (17/319) high school students; 53% (9/17) male; race/ethnicity: 35% (6/17) white, 18% (3/17) Latino, 29% (5/17) African American, 6% (1/17) Native American, and 12% (2/17) Asian	1
Public library	12-15	Overall, 2.2% (7/319) students; 100% (7/7) male and white	1
Youth program	13-18	Not reported	3
University undergraduate students	20-21	Not reported	3
University undergraduate and professional students	17-24	Overall, 5.0% (16/319); race/ethnicity: 44% (7/16) white and 56% (9/16) Asian	2
University student pharmacists	24-26	Overall, 17.2% (55/319); race/ethnicity: 91% (50/55) white	1
University science fair	Not reported	Children and parents (not reported)	N/A[a]

[a]Not applicable.

Community-Based Feedback

The questionnaire comprised open-ended questions about postgame opioid safety knowledge, experiences in gameplay, barriers and facilitators in gameplay, and opinions about the game. Group discussions after gameplay provided a mechanism for the project team to further understand gameplay barriers, facilitators, and overall feedback on the prototype. Responses were analyzed using content and thematic analysis involving open and axial coding. A total of 2 predominant themes emerged from the thematic analyses; the participants desired additional directions to guide gameplay and appreciated how the game demonstrated real-world scenarios and provided education that they could apply in real life.

The responses obtained from the questionnaire indicated that 78.4% (250/319) of the students perceived game scenarios as realistic. Participants reported the desire for more in-game hints and directions for gameplay. Participants reported that the game was appropriate for their age, whereas some high schoolers and college students thought that it was designed for younger populations. Overall, 94.6% (302/319) of the participants reported playing video games at home on a computer or phone every day. The scenarios, situations, and learning outcomes were reported to be realistic throughout the game. Suggestions for improvement from responses to the open-ended questionnaire included a need for additional background information, instructions, directions, or tutorials; expected actions; reducing the length of the dialog; increasing the speed of the dialog during gameplay; and updating the characters to reflect more human features. The group discussions shed light on the difficulties experienced by the players in navigating and transitioning between game scenarios and levels. The participants provided feedback on the game scenarios recommending that the safe and responsible medication management depicted in the game should align with the typical practices of families in both rural and urban settings.

Discussion

Principal Findings

The project used the serious game behavior change framework for improving medication safety in community settings to develop a serious game prototype, MEDSMAR$_x$T: Adventures in PharmaCity. The game prototype is targeted for use among adolescents to prevent prescription opioid misuse. The creation of the initial game prototype involved a methodological approach that was multidisciplinary, iterative, and participatory and involved community stakeholders [54]. The project team garnered preliminary feedback from youth and suggestions to improve the quality of gameplay and the prototype design. Overall, the game was perceived as realistic, and learning outcomes seemed achievable. The game prototype was assessed using feedback from middle school, high school, and college students through informal group discussions and surveys. Half of the respondents perceived that the game was suited to their age, and the scenarios were realistic. Respondents reported barriers in gameplay, such as limited instructions and technical difficulty with some gameplay features. The result from this formative project will further improve the MEDSMAR$_x$T: Adventures in PharmaCity game prototype. The feedback provides insights into the perceptions teens have about the game prototype. Although the game was perceived as realistic and the outcomes seemed achievable, teens also reported a need for iterations to the MEDSMAR$_x$T: Adventures in PharmaCity prototype. Future directions and several limiting factors of game development are discussed below.

Future Directions

The next step in this project is to polish and refine the MEDSMAR$_x$T: Adventures in PharmaCity game based on feedback from AYAs, conduct formal pilot testing of the revised prototype, and create an in-game analytics infrastructure to accommodate data collection from a large number of game users. Building this infrastructure will help determine if the domains of the conceptual framework were successfully met and the avenues where the game requires revisions based on internal gameplay data. In-game data such as time to complete a level, time spent at various decision-making points, correct and incorrect answers, and progress within each level may enable meaningful analysis to determine the effectiveness of promoting prescription opioid safety among AYAs. Formal testing of the game's effectiveness that incorporates in-game learning analytics data can provide a rich picture of a player's gameplay experience and the potential impact of the game. Testing this game will begin with a randomized control trial that would include building learning analytics infrastructure into the game. Adapting for smartphones will be done simultaneously as well.

Limiting Factors

Costs

Developing a serious game usually requires a much smaller budget than developing higher-quality commercial games. However, despite the availability of several game development and authoring tools, the complexity of developing a high-quality educational serious game is still beyond the reach of many developers. Creating a serious game incurs immense costs for game developers, especially for creating artistic assets such as graphics, characters, animation, dialog, and instructional design for players. The limited availability of funds prompted the project team to initiate the game design with key features, with future plans to incorporate in-depth gameplay infrastructure for data tracking for future iterations of the game.

Adoption and Deployment

Game deployment is the process of making the game available to the audience. Traditional PC games need to be installed in each machine before play. However, with current technology, providing a website to play the game is more common. The game must also be compatible with the operating system (ie, Windows or macOS) on the player's computer. With institutional policies regarding data safety and monitoring, setting up a game in classrooms, hospitals, or pharmacies can be a daunting task for teachers, hospitals, and pharmacy staff. The current prominent use of mobile devices and the high prevalence of cell phone usage among teenagers makes the situation more practical as the operating systems are less variable compared with computer systems.

Game adoption is the process of actually using the serious game in the desired audience to make learning through the game possible. The audience for game adoption, in this case, is schools and health care settings, which often report a lack of adequate training on how to integrate the serious game in their teaching or consultations. Proactively identifying strategies to promote adoption in community settings would require collaboration with relevant stakeholders such as educators and health care providers to assess the best practices for dissemination and implementation in their respective practices. In addition, opportunities to creatively integrate the serious game into health and science classrooms and medication counseling sessions in health care settings are important considerations. The process of enabling effective adoption of the game requires more time, resources, and iterative playtesting, which involves integrating the game into real practice to reach the desired audience. Involving educators and health care professionals is critical to promote adoption and dissemination in community settings.

Building Learning Analytics Infrastructure

A major next step is to develop the learning analytics for the MEDSMAR$_x$T: Adventures in PharmaCity game prototype. One study described learning analytics infrastructure as data collected in learning management systems (LMSs) or content management systems for improving teaching and learning [53]. One of the key aspects of learning analytics is tracking how many players interact with a web-based LMS by consulting access logs from the game server. An in-depth analysis of these logs can help identify behavior patterns that correlate with how the game is played. Incorporating the LMS features in game design such that it can be used for tracking gameplay requires financial resources and the expertise of game developers. In addition, this will enable data mining to extract and analyze these data. Involving all these components and experts in game design and development requires time, resources, and further playtesting.

Fast-Changing Technology

Rapidly changing and evolving technology is another challenge in disseminating this game to various audiences. Platforms for gameplay are constantly evolving, and adapting the game to the ever-evolving technology is both costly and time-consuming. This game prototype is currently developed for a computer but can easily be adapted to a smartphone in the future. Although challenging, keeping up with the constantly changing versions of smartphone and computer software would also be essential to overall game adoption and dissemination in the community.

Proposed Framework

The proposed framework is broad and has many categories. Further research and assessment are needed to understand the most important factors of the framework that directly affect gameplay, medication safety, and adolescent learning, which should be evaluated in quantitative trials [54].

Conclusions

In summary, developing an audience-centered game is essential for its mainstream adoption. Therefore, it is crucial that the game content, learning outcomes, objectives, and game complexity incorporate feedback from the targeted AYA audience [55]. The project team sought to obtain initial feedback on the current game prototype from many AYAs in community settings. The feedback on gameplay, game design, and technical issues assisted the project team in fine-tuning the MEDSMAR$_x$T: Adventures in PharmaCity game to ensure the creation of a consumer-centered prototype. The information learned through the extensive game prototype development process is foundational for improving and implementing MEDSMAR$_x$T: Adventures in PharmaCity on a larger scale in community settings.

This formative project describes the development process of MEDSMAR$_x$T: Adventures in PharmaCity. We propose a conceptual model entitled the *serious game behavior change framework for improving medication safety in community settings* grounded in social behavior and game development theories. The framework may be used for developing serious games to educate youth about prescription opioid safety**.** The project utilized a theory-driven conceptual framework for the development of a serious game targeting adolescent opioid misuse prevention and garnered preliminary feedback on the game to improve the quality of gameplay and the prototype. Feedback through informal assessments in community settings suggest that the youth and their families were interested in a game-based approach to learning about prescription opioid safety in homes and schools. The next steps include modifications to the game prototype based on community-based feedback, integrating learning analytics to track players' in-game behaviors, and formal testing of the final prototype. The future directions constitute making this game prototype more consumer-centered and integrating the learning analytics framework in the game to track the progress of players and gaming behaviors in gameplay.

Acknowledgments

The authors would like to thank Lisa Szela, Laura Stephenson, PhD, and Megan Buechel from the University of Wisconsin-Madison School of Pharmacy for their assistance in editing the manuscript. The authors would also like to thank the game development team from Gear Learning studio at the Wisconsin Center for Education Research, University of Wisconsin-Madison School of Education.

This study was supported by KL2 grant KL2 TR002374-03 and grant UL1TR002373 to University of Wisconsin-Madison Institute for Clinal and Translational Research by the Clinical and Translational Science Award program, through the National Institutes of Health (NIH) National Center for Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. Abraham O, Thakur T, Brown R. Prescription opioid misuse and the need to promote medication safety among adolescents. Res Social Adm Pharm 2019 Jul;15(7):841-844. [doi: 10.1016/j.sapharm.2019.01.003] [Medline: 30655180]
2. Prescription Opioids. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/drugoverdose/opioids/prescribed.html [accessed 2018-01-18]
3. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm [accessed 2018-01-11]
4. Costello M, Thompson S, Aurelien J, Luc T. Patient opioid education: research shows nurses' knowledge of opioids makes a difference. Medsurg Nurs 2016;25(5):307-311,333 [FREE Full text]
5. Benyamin R, Trescot AM, Datta S, Buenaventura R, Adlaka R, Sehgal N, et al. Opioid complications and side effects. Pain Physician 2008 Mar;11(2 Suppl):S105-S120 [FREE Full text] [Medline: 18443635]
6. Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings. SAMHSA - Substance Abuse and Mental Health Services. 2010. URL: https://www.samhsa.gov/data/sites/default/files/NSDUHNationalFindingsResults2010-web/2k10ResultsRev/NSDUHresultsRev2010.htm [accessed 2020-02-03]
7. de la Cruz M, Reddy A, Balankari V, Epner M, Frisbee-Hume S, Wu J, et al. The impact of an educational program on patient practices for safe use, storage, and disposal of opioids at a comprehensive cancer center. Oncologist 2017 Jan;22(1):115-121 [FREE Full text] [doi: 10.1634/theoncologist.2016-0266] [Medline: 27742907]

8. McCarthy DM, Wolf MS, McConnell R, Sears J, Chevrier A, Ahlstrom E, et al. Improving patient knowledge and safe use of opioids: a randomized controlled trial. Acad Emerg Med 2015 Mar;22(3):331-339 [FREE Full text] [doi: 10.1111/acem.12600] [Medline: 25731073]
9. Chakravarthy B, Somasundaram S, Mogi J, Burns R, Hoonpongsimanont W, Wiechmann W, et al. Randomized pilot trial measuring knowledge acquisition of opioid education in emergency department patients using a novel media platform. Subst Abus 2018 Jan 2;39(1):27-31. [doi: 10.1080/08897077.2017.1375061] [Medline: 28873050]
10. Williams MV, Parker RM, Baker DW, Parikh NS, Pitkin K, Coates WC, et al. Inadequate functional health literacy among patients at two public hospitals. J Am Med Assoc 1995 Dec 6;274(21):1677-1682. [Medline: 7474271]
11. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]
12. Howard DH, Gazmararian J, Parker RM. The impact of low health literacy on the medical costs of medicare managed care enrollees. Am J Med 2005 Apr;118(4):371-377. [doi: 10.1016/j.amjmed.2005.01.010] [Medline: 15808134]
13. Williams DM, Counselman FL, Caggiano CD. Emergency department discharge instructions and patient literacy: a problem of disparity. Am J Emerg Med 1996 Jan;14(1):19-22. [doi: 10.1016/S0735-6757(96)90006-6] [Medline: 8630148]
14. Powers RD. Emergency department patient literacy and the readability of patient-directed materials. Ann Emerg Med 1988 Feb;17(2):124-126. [doi: 10.1016/s0196-0644(88)80295-6] [Medline: 3337429]
15. Rampey B, Finnegan R, Goodman M. Skills of US Unemployed, Young, and Older Adults in Sharper Focus: Results From the Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014: First Look. The National Center for Education Statistics (NCES). 2016. URL: https://nces.ed.gov/pubs2016/2016039.pdf [accessed 2020-05-29]
16. Dunn KM, Saunders KW, Rutter CM, Banta-Green CJ, Merrill JO, Sullivan MD, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med 2010 Jan 19;152(2):85-92 [FREE Full text] [doi: 10.7326/0003-4819-152-2-201001190-00006] [Medline: 20083827]
17. Jones CM, Mack KA, Paulozzi LJ. Pharmaceutical overdose deaths, United States, 2010. J Am Med Assoc 2013 Feb 20;309(7):657-659. [doi: 10.1001/jama.2013.272] [Medline: 23423407]
18. Kuehn BM. Efforts aim to curb opioid deaths, injuries. J Am Med Assoc 2009 Mar 25;301(12):1213-1215. [doi: 10.1001/jama.2009.367] [Medline: 19318643]
19. Schwartzberg J, VanGeest J, Wang C. Understanding Health Literacy: Implications For Medicine And Public Health. Chicago, IL: American Medical Association Press; 2004.
20. Kutner M, Greenberg E, Jin Y, Paulsen C. The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy. The National Center for Education Statistics (NCES). 2004. URL: http://nces.ed.gov/pubs2006/2006483.pdf [accessed 2020-02-01]
21. Girard C, Ecalle J, Magnan A. Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. J Comput Assist Learn 2012 Jun 13;29(3):207-219 [FREE Full text] [doi: 10.1111/j.1365-2729.2012.00489.x]
22. Bruggers CS, Altizer RA, Kessler RR, Caldwell CB, Coppersmith K, Warner L, et al. Patient-empowerment interactive technologies. Sci Transl Med 2012 Sep 19;4(152):152ps16. [doi: 10.1126/scitranslmed.3004009] [Medline: 22993292]
23. Susi T, Johannesson M, Backlund P. Serious Games – An Overview. DiVA Portal. 2007. URL: https://www.diva-portal.org/smash/get/diva2:2416/FULLTEXT01.pdf [accessed 2020-05-29]
24. Marsh T. Serious games continuum: between games for purpose and experiential environments for purpose. Entertain Comput 2011 Jan;2(2):61-68 [FREE Full text] [doi: 10.1016/j.entcom.2010.12.004]
25. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for health for children-current status and needed research. Games Health J 2016 Feb;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
26. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22(1):77-89. [doi: 10.3109/14639239709089835] [Medline: 9183781]
27. Horan PP, Yarborough MC, Besigel G, Carlson DR. Computer-assisted self-control of diabetes by adolescents. Diabetes Educ 1990;16(3):205-211. [doi: 10.1177/014572179001600311] [Medline: 2185006]
28. Krishna S, Francisco BD, Balas EA, König P, Graff GR, Madsen RW, Randomized Trial. Internet-enabled interactive multimedia asthma education program: a randomized trial. Pediatrics 2003 Mar;111(3):503-510. [doi: 10.1542/peds.111.3.503] [Medline: 12612228]
29. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature reviewInvestigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020 Apr 29;8(2):e16096 [FREE Full text] [doi: 10.2196/16096] [Medline: 32347811]
30. Whiteley L, Mena L, Craker LK, Healy MG, Brown LK. Creating a theoretically grounded gaming app to increase adherence to pre-exposure prophylaxis: lessons from the development of the viral combat mobile phone game. JMIR Serious Games 2019 Mar 27;7(1):e11861 [FREE Full text] [doi: 10.2196/11861] [Medline: 30916652]
31. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018;30(sup4):21-33 [FREE Full text] [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]

32. Abraham O, Feathers A, Grieve L, Babichenko D. Developing and piloting a serious game to educate children about over - the - counter medication safety. J Pharm Health Serv Res 2019 Feb 21;10(2):235-241 [FREE Full text] [doi: 10.1111/jphs.12292]
33. Castel AD, Qasmieh S, Greenberg D, Ellenberger N, Howell TH, Griffith C, et al. Digital gaming to improve adherence among adolescents and young adults living with HIV: mixed-methods study to test feasibility and acceptability. JMIR Serious Games 2018 Oct 15;6(4):e10213 [FREE Full text] [doi: 10.2196/10213] [Medline: 30322838]
34. de Smet A, van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
35. Thompson D, Baranowski T, Buday R, Baranowski J, Thompson V, Jago R, et al. Serious video games for health how behavioral science guided the development of a serious video game. Simul Gaming 2010 Aug 1;41(4):587-606 [FREE Full text] [doi: 10.1177/1046878108328087] [Medline: 20711522]
36. Duncan LR, Hieftje KD, Culyba S, Fiellin LE. Game playbooks: tools to guide multidisciplinary teams in developing videogame-based behavior change interventions. Transl Behav Med 2014 Mar;4(1):108-116 [FREE Full text] [doi: 10.1007/s13142-013-0246-8] [Medline: 24653781]
37. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
38. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, UK: Prentice Hall; 1986.
39. Gee JP. What video games have to teach us about learning and literacy. Comput Entertain 2003 Oct 1;1(1):20. [doi: 10.1145/950566.950595]
40. Schank R, Schank RC. Designing World-Class E-Learning : How IBM, GE, Harvard Business School, And Columbia University Are Succeeding At E-Learning. New York, USA: McGraw-Hill; 2002.
41. Schank R, Fano A, Bell B, Jona M. The design of goal-based scenarios. J Learn Sci 1994 Oct;3(4):305-345 [FREE Full text] [doi: 10.1207/s15327809jls0304_2]
42. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for ehealth and mhealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
43. Thalheimer W. Using Linguistically, Culturally, and Situationally Appropriate Scenarios to Support Real-World Remembering. Work-Learning Research. URL: http://www.work-learning.com/catalog/ [accessed 2018-08-01]
44. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
45. Prince M. Does active learning work? A review of the research. J Eng Educ 2004;93:223-231. [doi: 10.1002/j.2168-9830.2004.tb00809.x]
46. Thomas RC, Weywadt CR, Anderson JL, Martinez-Papponi B, McDaniel MA. Testing encourages transfer between factual and application questions in an online learning environment. J Appl Res Mem Cogn 2018 Jun;7(2):252-260. [doi: 10.1016/j.jarmac.2018.03.007]
47. Kapur M, Bielaczyc K. Designing for productive failure. J Learn Sci 2012 Jan;21(1):45-83. [doi: 10.1080/10508406.2011.591717]
48. Generation RX and Cardinal Health. 2019. URL: https://www.generationrx.org/ [accessed 2020-01-17]
49. Dose of Reality. 2020. URL: https://doseofrealitywi.gov/ [accessed 2020-01-17]
50. Chronic Pain Community- Stories on Managing Pain. 2020. URL: http://www.yourpaincommunity.com/ [accessed 2020-01-17]
51. Nittono H, Fukushima M, Yano A, Moriya H. The power of Kawaii: viewing cute images promotes a careful behavior and narrows attentional focus. PLoS One 2012;7(9):e46362 [FREE Full text] [doi: 10.1371/journal.pone.0046362] [Medline: 23050022]
52. Alaswad Z, Nadolny L. Designing for game-based learning: the effective integration of technology to support learning. J Educ Technol Syst 2015 Jul 17;43(4):389-402 [FREE Full text] [doi: 10.1177/0047239515588164]
53. Elias T. Learning Analytics: Definitions, Processes and Potential. Semantic Scholar. 2011. URL: https://pdfs.semanticscholar.org/732e/452659685fe3950b0e515a28ce89d9c5592a.pdf [accessed 2020-01-04]
54. Verschueren S, Buffel C, Stichele GV. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019 May 2;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
55. Quintana C, Shin N, Norris C, Soloway E. Learner-centered design. In: Sawyer RK, editor. The Cambridge Handbook of the Learning Sciences. Cambridge, UK: Cambridge University Press; 2006.

Abbreviations

AYA: adolescent and emerging young adult

LMS: learning management system
NIH: National Institutes of Health

Edited by G Eysenbach; submitted 11.02.20; peer-reviewed by A Goedken, L Whiteley; comments to author 16.03.20; revised version received 08.05.20; accepted 13.05.20; published 03.07.20.

Please cite as:
Abraham O, Thakur T, Brown R
Developing a Theory-Driven Serious Game to Promote Prescription Opioid Safety Among Adolescents: Mixed Methods Study
JMIR Serious Games 2020;8(3):e18207
URL: https://games.jmir.org/2020/3/e18207
doi:10.2196/18207
PMID:32618571

Review

Mapping Behavioral Health Serious Game Interventions for Adults With Chronic Illness: Scoping Review

Teresa Hagan Thomas[1], PhD, RN; Varshini Sivakumar[1], SN; Dmitriy Babichenko[2], PhD; Victoria L B Grieve[3], PharmD; Mary Lou Klem[4], PhD, MLIS

[1]School of Nursing, University of Pittsburgh, Pittsburgh, PA, United States
[2]School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, United States
[3]Department of Pharmacy and Therapeutics, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States
[4]Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Teresa Hagan Thomas, PhD, RN
School of Nursing
University of Pittsburgh
440 Victoria Building
3500 Victoria Street
Pittsburgh, PA, 15261
United States
Phone: 1 4126243799
Email: t.thomas@pitt.edu

Abstract

Background: Serious games for health are increasingly being used to address health outcomes in patients with chronic illnesses. These studies vary in their study designs, patient populations, frameworks, outcome variables, and degree of specificity of the serious game intervention.

Objective: This scoping review aims to clarify the conceptual features of the existing research related to serious games designed to improve cognitive and behavioral outcomes in adults with chronic illness.

Methods: We applied the Preferred Reporting Items of Systematic Reviews and Meta-Analysis for scoping reviews (PRISMA-ScR) methodology, including an a priori research question. We searched 4 electronic databases to identify articles published through November 2019. Inclusion criteria encompassed (1) adults 18 years or older; (2) patients with a diagnosis of chronic illness; (3) a serious game intervention; and (4) defined patient outcomes that assess patients' behavioral, cognitive, or health outcomes.

Results: Of the 3305 articles identified, 38 were included in the review. We charted and analyzed the theoretical frameworks, key concepts, and outcome variables of these studies with summaries of features across articles. The majority of studies used a randomized controlled trial design (23/38, 61%), included a custom serious game intervention (22/38, 58%), and lacked a theoretical framework (25/38, 66%). Common outcome variables included quality of life (16/38, 42%), mood (15/38, 39%), cognitive function (13/38, 34%), symptoms (12/38, 32%), and physical activity (9/38, 24%). Key differences between studies included whether or not serious games aimed to train versus teach patients, be widely accessible versus tailored interventions, or replace versus complement current treatments.

Conclusions: This scoping review defines the current landscape of research in serious games for health research targeting behavioral and cognitive outcomes in adults with chronic disease. Studies have addressed a variety of patient populations and diverse patient outcomes. Researchers wanting to build on the current research should integrate theoretical frameworks into the design of the intervention and trial to more clearly articulate the active ingredients and mechanisms of serious games.

(JMIR Serious Games 2020;8(3):e18687) doi:10.2196/18687

KEYWORDS

review; chronic disease; behavioral sciences; video games

Introduction

Background

Over the past decade, health-oriented clinical and research apps using electronic serious games have increased as a means to improve patient outcomes and provide health education [1]. Serious games take important health topics traditionally taught to patients and apply game features to provide a motivational, engaging, and even fun learning experience [2]. These games may be used to prevent disease [3], improve the health of patients with disease [4], and enhance social interaction to improve health [5]. Additional genres of serious games teach medical professionals skills [6-8], ultimately attempting to improve patient outcomes by improving clinician knowledge. A central attraction of electronic serious games is their ability to educate, motivate, and involve users without using conventional patient education that relies heavily on less engaging and often more intense training, such as written instruction or one-on-one consultation. The potential benefits of serious games include their high level of learning engagement, ease, and low cost of dissemination and distribution [9].

McGonigal [10], a game designer and leader in developing games to improve quality of life, defined serious games as having 4 hallmark features: (1) an *overall goal*—some desired outcome that provides a sense of purpose, (2) *rules*—limitations on the users' activities that necessitate creativity and strategic thinking, (3) *feedback system*—a way to communicate with users about their proximity to achieving a goal that motivates and promises that the goal is achievable, and (4) *voluntary participation*—the freedom to enter and leave the game so that participation is safe and pleasurable. The theorized learning mechanism leading serious games to be an effective teaching tool involves immersive qualities in which the users becomes engrossed in the game; the requirement for users to learn skills in increasingly difficult challenges; and use of the user's desire for mastery, arousal, diversion, and challenge [10,11].

Transformational games have emerged as a subset of serious games that try to positively impact the user by addressing outcomes, including behaviors, attitudes, and social issues [12]. Unlike traditional educational games that focus on the game as an end in itself, transformational games aim to have users learn through intentional participation in narratives that employ concepts that, if successfully learned, should extend beyond the game and meaningfully impact their lives [13]. This generalization requires that the game have specific behavior and learning outcomes, including an explicit plan for how users will transfer their skills from the game into real-world settings.

Rehabilitation and behavioral sciences have pioneered novel motivational, engaging games. Currently, the field is rapidly expanding to new patient populations and topics. Patients with chronic illnesses such as cancer, cardiovascular disease, diabetes, and obesity face long-term health problems that require diligent management. For these patients, behavioral, cognitive, and health outcomes are essential in ensuring that they can self-manage their illness and prevent long-term morbidity and mortality. These patient outcomes cut across different chronic diseases and include physical activity, maintenance of a healthy body weight, quality of life, symptom burden, mood, and cognitive function.

As electronic serious game interventions for health have increased, a lack of attention to the theoretical underpinnings has resulted in disparate game mechanisms, unclear theoretical frameworks of action, and weak study designs. Previous review articles have noted a need for improved research methods and expanded clinical applicability [14-17]. A fresh review of serious games in adults with chronic illnesses can assist researchers and clinicians in recognizing the strengths and limitations of how these studies have been designed. AA review can also identify ways to improve research rigor and translation to improved patient outcomes. Mapping the current research landscape will assist in establishing the key concepts, variables, theories, and frameworks undergirding this growing body of research.

Objective

Scoping reviews allow investigators to systematically examine emerging areas of evidence and can help identify gaps in knowledge, clarify concepts, and reveal methodological concerns in new areas of research. As scoping reviews generally have broader inclusion criteria than systematic reviews without an assessment of study quality, they allow for findings from disparate patient populations and contexts and a more comprehensive determination of evidence. Researchers and clinicians can apply the findings of a scoping review to more astutely build on the current evidence base and address gaps in existing research. Although systematic reviews of serious games have been conducted within specific patient populations and health care settings [18-21], the evidence supporting serious games focused on health skills and behaviors among adults with chronic diseases have not been mapped systematically. This scoping review aimed to define the concepts applied to studies using serious games to improve the health of adults with chronic illnesses.

Methods

Research Question

Our *a priori* research question was as follows: "What types of theoretical frameworks, key concepts, and outcome variables exist within serious game interventions to improve the cognitive and behavioral outcomes of adult patients with chronic illness?"

Protocol

We followed a scoping review methodology to synthesize concepts and research concerning the use of serious games as interventions designed to address cognitive and behavioral outcomes among adult patients with chronic illness. The objectives, inclusion criteria, and methods for this scoping review were specified in advance and documented in a protocol. Our protocol is freely available through the Open Science Framework [22]. This protocol follows the Preferred Reporting Items of Systematic Reviews and Meta-Analysis for scoping reviews (PRISMA-ScR) methodology to ensure that our results would be systematically conducted with minimal bias [23].

Inclusion Criteria

To be included in the scoping review, articles needed to focus on the following: (1) adults 18 years or older; (2) patients with a diagnosis of chronic illness; (3) a serious game intervention; and (4) defined patient outcomes that assess patients' behavioral, cognitive, and health outcomes.

Participants

We used definitions of chronic illness as defined by the Centers for Disease Control and Prevention [24] or the Center for Medicare and Medicaid Services [25]. These illnesses include Alzheimer's disease, arthritis, cancer, cardiovascular diseases including hypertension and stroke, chronic lung disease including asthma and chronic obstructive pulmonary disease (COPD), cystic fibrosis, diabetes, epilepsy and seizures, obesity, and oral health. Some articles included participants with illnesses that require long-term maintenance and ongoing medical care but were not listed in these definitions [26]. To err on the side of inclusion, we revised our definition of chronic illness to include additional diseases that could be considered chronic, such as Parkinson's disease. When studies included participants with and without chronic illnesses, we reviewed the full-text article to determine the proportion of participants with chronic illnesses and only included articles in which the majority of participants had an eligible disease.

Concept

We broadly defined serious games as "games that are designed to entertain players as they educate, train, or change behavior" [27]. We used the 4 criteria by McGonigal [10] listed above to determine whether an intervention included the requisite elements of a serious game. We did not stipulate that the serious game required any specified dose or intensity, length, or use of a comparison group. When limited descriptions of the game were available, we made assumptions about the existence of certain features to be liberal in article selection. Some articles included screenshots of the serious game, which indicated that the game included feedback, points, and engaging characters and scenery despite the article not describing these features. We did not include articles for which the serious game was not the intervention but was conceptualized as a diagnostic tool or priming event for a separate intervention. For example, some studies used video games to prime study patients' working memory and cognitive abilities, but the study was not designed to test the effect of that video game on patients' outcomes.

Outcomes

We included any study that assessed patients' behavioral, cognitive, and health outcomes. We created an ongoing list of outcome variables based on the variables identified in specific studies. Examples of behavioral outcomes included physical activity, medication adherence, and self-management.Examples of cognitive outcomes included executive function, working memory, learning new knowledge, and motivation. Examples of health outcomes included quality of life, mood, symptoms, BMI, hemoglobin A_{1c} (HbA_{1c}), and blood pressure. We did not include functional outcomes (eg, the impact of a serious game on hand function or range of motion) because these assessments did not include the patient's involvement in learning or behaving to manage their chronic disease.

Context

As many interventions and meta-analyses have studied the impact of games to increase patients' mobility and functional outcomes [28] and do not include behavioral or cognitive outcomes, we only included serious game interventions that included behavioral, cognitive, or related health outcomes. For example, a study that analyzed the effect of a serious game on patients' motor control would not be included if the main outcomes only included functional measurements. A study would be included if the main outcomes included measures of patient self-management of their illness or symptoms.

Identifying Relevant Studies and Study Selection

The following databases were searched: PubMed (1946 to November 28, 2017), EbscoHost CINAHL (Cumulative Index of Nursing and Allied Health Literature; 1981 to November 2017), Ovid PsycINFO (1967 to November, Week 2, 2017), and EMBASE (Excerpta Medica dataBASE; 1974 to November 28, 2017). All database searches were run on November 28, 2017, and, when available, a search limit to the English language was applied. Updated searches were run on November 25, 2019. An experienced health sciences librarian (MK) designed the PubMed search [22] and then translated that search for use in the other databases. The search strings consisted of natural language terms and (when available) controlled vocabulary representing the concepts of "videogames," "serious games," and "chronic diseases."

Two individuals independently screened abstracts and full-text articles using DistillerSR software (DistillerSR, Evidence Partners). This software allowed reviewers to collaboratively review all abstracts for the inclusion criteria. For all included articles, 2 reviewers extracted data and agreed upon the results. When reviewers disagreed on the inclusion criteria or data, they met to discuss the articles until consensus was reached.

Articles that met the inclusion criteria were charted to provide a descriptive summary. We extracted data including their study design; patient population; frameworks; behavioral, cognitive, or health-related outcomes; and a description of the serious game. We further described the key concepts that emerged as similarities and differences across the studies.

Results

The search strategy yielded 3268 references, of which 40 met the inclusion criteria. Figure 1 illustrates the PRISMA-ScR flowchart for our results. One research study had 2 manuscripts reporting different results from the same trial. As the framework, sample, and intervention were the same despite reporting on different outcome variables, we grouped these manuscripts as one study. A separate group of investigators had 2 manuscripts for the same trial and a separate manuscript reporting the study protocol. We grouped these manuscripts together, prioritizing the trial results. Our final sample included 38 studies. Table 1 illustrates the frequency of trial designs, populations, and most frequently cited study outcomes.

Figure 1. Preferred Reporting Items of Systematic Reviews and Meta-Analysis for scoping reviews flowchart.

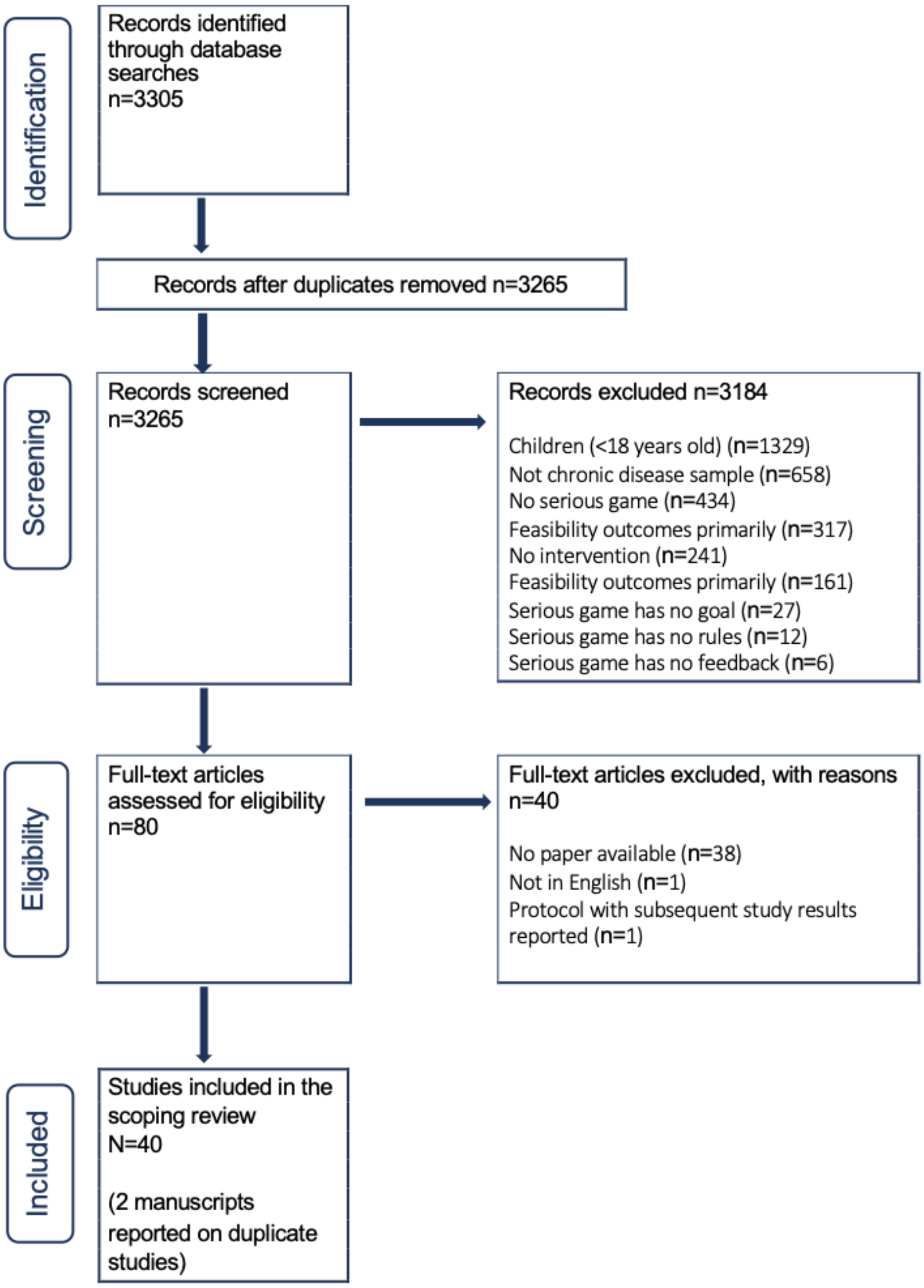


Table 1. Frequency of trial designs, patient populations, and study outcome variables.

Characteristics	Values
Trial design, n (%)	
Randomized controlled trial	23 (61)
Feasibility study	8 (21)
Protocol for study	6 (16)
Case study	1 (3)
Patient population, n (%)	
Stroke	17 (45)
Cancer	6 (16)
Diabetes	5 (13)
Dementia	3 (8)
Hypertension	3 (8)
Parkinson's disease	2 (5)
Obesity	1 (3)
Chronic obstructive pulmonary disease	1 (3)
Outcome variable, n (%)	
Quality of life	16 (42)
Mood	15 (39)
Cognitive function	13 (34)
Symptoms	12 (32)
Physical activity	9 (24)

Trial Designs

The majority of studies (23/38, 61%) reported the results of randomized controlled trials, whereas 16% (6/38) articles were protocols for clinical trials. Of these articles, the comparison groups for these trials mostly included a conventional form of active therapy (23/29, 79%), such as rehabilitation, rather than usual care or attention control (6/29, 21%). Few studies in the stroke population used a usual care comparison group, opting for conventional rehabilitation as a comparison. Less frequent trial designs included one-group pretest/posttest feasibility evaluations (8/38, 21%) and a case study (1/38, 3%).

Populations

The patient populations represented a diverse group of common chronic diseases, including stroke (17/38, 45%), cancer (6/38, 16%), diabetes (5/38, 13%), dementia (3/38, 8%), hypertension (3/38, 8%), Parkinson's disease (2/38, 5%), obesity (1/38, 3%), and COPD (1/38, 3%). The mean sample size was 82 patients (SD 104), with a range of 1 to 456. Most studies had small samples that were not adequately powered to test for statistically significant differences.

Serious Games

The majority of studies (22/38, 58%) described serious games that were custom-made interventions for specific patient populations. Studies describing custom-made games tended to have an in-depth description of the game's features, often because the research team built the game before evaluating it. Three custom-made games used a version of the BrightArm virtual reality system [29-31]; all other games were evaluated in separate manuscripts.

The remaining studies (16/38, 42%) used off-the-shelf games designed for the general population. Off-the-shelf games included games within the Nintendo Wii Fit (11/38, 28%), Xbox Kinect (3/38, 8%), and other off-the-shelf software (2/38, 5%). Often, the authors did not explicitly state the features of the off-the-shelf games. In lieu of such detail, we assumed that games such as bowling, tennis, and ping pong had a goal (score the most points), restrictions on actions (only specific actions are allowed), and feedback (points, badges, and scoreboards). Although these off-the-shelf games are commonly used as exercise, physical activity, and mobility interventions, different studies specified different rationales for the same technology. For example, different studies used the Nintendo Wii Fit game suite to improve patient motor function, balance, cognition, or general physical activity. Of the 17 studies reporting on patients recovering from a stroke, 8 (47%) used off-the-shelf games.

Theoretical Frameworks

Most of the articles (25/38, 66%) lacked a clear theoretical framework. This was especially true for articles focused on patients with a history of stroke; only 3 of the 17 studies included a theoretical framework [31-33]. In most cases, the behavioral or cognitive outcomes appeared to serve as an *add-on* to other functional outcomes without an underlying hypothesis proposing how the serious game leads to improved outcomes.

For example, authors may include measures of quality of life, depression, or motivation to their study but not indicate how these health outcomes related to the intervention.

Within the 34% (13/38) studies that cited a theoretical framework, frameworks varied in their application to the intervention and variables. Most frameworks focused on mechanics of specific interventions (simulation theory [33], cognitive training reorganization on brain network infrastructure [34], spaced education [35], neurofeedback for pain control [36], integrative rehabilitation [31], and self-management [33,37]). Other frameworks focused on intervention development (intervention mapping approach [38] and Gagne's instruction strategies [39]) and rationales for serious game features (behavioral economics [40,41], dual-task training [32], narrative transportation theory [42], self-determination theory [42-44], and behavior change theory [43,44]).

Outcome Variables

Studies varied greatly in the type and number of outcome variables used to assess the efficacy of the serious game. We report the most frequently cited outcome variables while noting several additional outcome variables. Multimedia Appendix 1 [29-68] includes a complete list of the behavioral, cognitive, and health outcomes as conceptualized within each study. We indicate whether the data for each outcome are based on patient self-report; examination with a valid assessment or clinical assessment; and/or objectively collected through sensor data, blood work, or medical records.

Quality of Life

Quality of life (including health-related quality of life) was the most common outcome across all studies (16/38; 42%), including a variety of chronic diseases. Many studies included quality of life as an outcome, seeming to want to capture a patient's overall well-being. The rationale for including quality of life was rarely described. It was frequently captured as a secondary outcome to functional outcomes of mobility and strength in rehabilitation serious games for patients recovering from a stroke. Quality of life measures included self-report surveys, including the EuroQol-5 Dimension, Short Form Survey-36, and Stroke Impact Scale.

Mood

Mood—including depression and anxiety—was included as an outcome in 39% (15/38) studies. Similar to quality of life, mood variables were often included with minimal description tying these outcomes to the use and intent of the serious game intervention. Studies frequently included the Beck Depression Inventory as an outcome of interest, although the rationale for including this scale was lacking, and therefore, the discussion of these findings was minimal. Other studies measured patients' self-reported anxiety using the State-Trait Anxiety Inventory or disease-specific mood surveys.

Cognitive Function

Many serious games have been designed to improve memory and attention in patients with a history of stroke, dementia, and Parkinson's disease. Cognitive function was a common outcome in studies (13/38, 34%), especially for serious games targeting patients with dementia and stroke. Cognitive function assessments commonly included the Trail Making Test, Neuropsychological Assessment Battery, Mini-Mental Status Examination, and other neuropsychological evaluations that were administered by a trained study team member.

Symptoms

Physical symptoms such as fatigue and pain were reported as outcomes of the serious game in 32% (12/38) studies. Specific symptoms varied widely and were typically tailored to the patient population within the article (eg, fatigue was commonly assessed for patients with cancer [38,42,49,50]). Symptom measures included the Functional Assessment of Cancer Therapy, McGill Pain Questionnaire, and other self-report measures. Some studies measured symptoms using patient medical records that indicated symptom ratings and medication use to manage symptoms.

Physical Activity

A total of 24% studies (9/38) included a measure of physical activity as an outcome variable. In this case, physical activity was considered an activity of diabetes self-management, which would lead to better control of diabetes. Physical activity measures ranged from passive sensor data from wearable devices (eg, pedometers, accelerometers, and sensors on phones), clinical examinations of patients' cardiorespiratory fitness, to self-reported activity surveys.

Other Outcomes

Studies assessed myriad additional outcomes. Several studies assessed patients' motivation (5/38, 13%) in engaging in a serious game, captured by various self-report measures. Several studies assessed physical measurements conceptualized as health outcomes of desired behaviors. For example, studies targeting patients with diabetes often collected blood levels of HbA_{1c} (4/38, 11%) as an outcome variable indicating whether or not patients improved their average blood sugar level. Similarly, 3% study (1/38) assessed blood pressure as a target health outcome for patients with hypertension. Additional outcomes included patient self-reports of education or knowledge, self-management, medication adherence, and other outcomes. A minority of studies included outcomes related to learning within a serious game, fidelity of the intervention, or transference of learning within the game to real-life settings.

Key Concepts

Teaching Versus Training

A significant difference existed between the serious games that were described as teaching patients a specific behavior or skill compared with games intended to train patients in movements or activities. Games that were focused on teaching tended to have more explicit descriptions tying the features of the serious game to the intended learning outcomes, whereas games that were focused on training tended to have minimal explanation of how the game would train patients. Several studies employing a teaching approach cited theoretical frameworks linking how learning a behavior would lead to improved outcomes. For example, Kerfoot et al [35] explicitly stated how a spaced education and self-management framework would teach patients how to manage their diabetes, thereby reducing their HbA_{1c} and diabetes distress and increasing their knowledge of diabetes.

Accessibility Versus Tailored

Differences between the off-the-shelf and custom-made serious games was a consistent concept discussed across the studies. Many studies cited the advantages of using an off-the-shelf serious game such as Nintendo Wii Fit or Xbox Kinect, including easy integration into community settings and dissemination of effective protocols. Alternatively, studies in which the serious game was tailored to a specific patient population or health problem discussed the advantages of integrating specific content and mechanisms known to be important and relevant to that patient population. Tailoring may include specific technologies required to address specific health problems or a more general consideration of the psychosocial needs of the population. The Memory Matters game by Yu et al [68] intricately designed their interactive reminiscence game for patients with dementia, providing objects, images, and music familiar to the target patient population.

Replacement Versus Complementary

Several studies described the rationale for serious game intervention as a replacement for conventional treatment. In many cases, serious games were seen as a substitute for occupational or rehabilitation therapy in patients recovering from a stroke. Shin et al [61] created the RAPAEL Smart Glove to simulate upper extremity rehabilitation and compared it with conventional occupational therapy. This varied vastly from serious games meant to provide additional, novel, and meaningful support but otherwise not replacing an existing intervention, such as interventions focused on teaching self-management or coping skills. For example, Höchsmann et al [43,44] reported on their intervention—MOBIGAME—which aimed to reduce diabetes by engaging patients in an immersive, relaxing program focused on increasing physical activity, motivation, and adherence.

Discussion

Primary Findings

This scoping review identified that electronic serious games for patients with chronic illness target a variety of patient populations, are mostly custom-made games, largely lack theoretical frameworks, and measure a broad array of patient outcomes. Common themes across studies included whether or not games were intended for teaching versus training purposes, meant to be widely available versus tailored to patient populations, and replace or complement existing therapies.

The 38 studies in this review represented 8 different patient populations, indicating that serious game interventions are applicable across diverse chronic illnesses. Many of the studies were designed for patients recovering from a stroke, possibly because off-the-shelf games (used in almost half of the studies for patients with stroke) included physical and functional targets similar to those used in conventional rehabilitation therapy. Nonetheless, a variety of serious games have been used across and within populations with chronic illness, underscoring the diversity of designs and settings in which researchers are investigating the efficacy of serious games.

A major limitation of the current landscape of serious games for adults with chronic illness is the overwhelming lack of theoretical frameworks. Two-thirds of studies did not cite a theoretical framework guiding the intervention, trial design, or proposed mechanisms linking the serious game to patient outcomes. This finding is similar to a recent systematic review of serious games [69]. One implication of this lack of frameworks is the broad list of study outcomes and lack of mechanistic or fidelity assessments. Outcomes were measured using diverse self-report measures, examinations, and sensor data, limiting the potential to compare outcomes across studies. The most frequently included outcomes, including quality of life and mood, were added without explicit hypotheses or proposed mechanisms linking them to the serious game. This finding corroborates other research [70], including a scoping review by Rohrbach et al [71], which found that virtual reality interventions for patients with stroke often include affective traits such as motivation and enjoyment without integration into the overall design and mechanistic planning of the intervention.

Recent consensus statements for the consistent integration of theoretical frameworks into serious games for health research provide researchers, developers, and clinicians with explicit recommendations on how to address this limitation of current studies [72]. Ideally, serious games for health would leverage the active ingredients of games (immersive, challenging, and chance for mastery) with persuasive strategies commonly used in behavior change research. Reassuringly, 5 of the 9 (56%) studies published in 2019 cited a theoretical framework, perhaps indicating a trend in serious game research. In the future, research should integrate gamification and behavior change theory into rigorously designed trials based on theoretical frameworks [73].

Few studies have assessed the transfer of skills learned within a serious game into real-life settings. Although some studies assessed patient behaviors to determine whether the serious game led to changes in lifestyle and activity, this was infrequent and sometimes only assessed during the intervention period. A systematic review by Kuipers et al [74] noted a lack of attention to the mechanistic underpinnings of how serious games facilitate learning and drive their intended long-term effects. The authors urge researchers to consider game transference effects in the design of serious games for health, increasing the likelihood—or at least the ability to document—that these health interventions will lead to lasting desired patient behaviors and outcomes. Only one study addressed how a serious game was integrated into patient care via an existing research and clinical platform [40]. Researchers designing custom serious games for health should include upfront plans for disseminating their intervention within clinical care or broadly into larger populations.

The outcome variables of the studies included were inconsistently aligned with the serious game's goals. For example, games designed to improve cognition and physical fitness measured outcomes of executive function and physical activity, respectively. On the other hand, many studies included outcomes that were less clearly aligned with the intervention.

This was especially true for studies that measured quality of life and mood. Many of these studies did not indicate a rationale for why quality of life or mood would change because of the serious game but appeared to be included as measures of patients' general well-being.

This scoping review has limitations. First, this review relied on published manuscripts of electronic serious games. We attempted to include a broad definition of both electronic serious games and chronic illnesses, although fluctuations in the terminology around these terms limited our search strategies' abilities to identify articles. Conference proceedings did not include sufficient detail to answer our research questions. Analog serious games were not included in this review and could be considered in future analyses. Second, this review relied on descriptions of serious games and interventions described within manuscripts, which were often limited, especially concerning game features and frameworks. Although we attempted to use our best judgment, additional details of the games and interventions could provide a more comprehensive summary. Finally, the broad nature of this review—although appropriate for a scoping review—limits the ability to provide specific conclusions based on patient population, game design, patient outcomes, etc. Future research could include more broader definitions of serious games, request information from authors to clarify game features and frameworks, and assess the efficacy of the serious game.

Conclusions

This review assists researchers in creating serious game interventions to address chronic health conditions by providing clarity on how to build from the current structure of serious games for health research. Researchers should continue the existing momentum in building robust, large trials driven by theoretical underpinnings of how interventions are hypothesized to impact outcomes. As attention to how game features lead to behavioral, cognitive, and health outcomes increases, we as a field will grow in the ability of our research to effectively and efficiently impact the most significant health problems patients experience.

Acknowledgments

This study was funded by the American Cancer Society Mentored Research Scholar Grant (MRSG-18-051-01) and the National Palliative Care Research Center Career Development Award.

Authors' Contributions

TT was the principal investigator. TT was involved in study conception, design, data capture, data analysis, and interpretation of the results and was the primary writer of the manuscript. VS was involved in data capture, data analysis, and editing of the manuscript. DB was involved in study conception, data analysis, and editing of the manuscript. VG was involved in study conception, data analysis, and editing of the manuscript. MK was involved in study conception, data capture, data analysis, and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results table of studies of serious games for adults with chronic disease targeting behavioral, cognitive, or other health outcomes (N=38).

[DOCX File , 35 KB - games_v8i3e18687_app1.docx]

References

1. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Khodabandeh ME, Heidari S, et al. Health education serious games targeting health care providers, patients, and public health users: scoping review. JMIR Serious Games 2020 Mar 5;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
2. Garris R, Ahlers R, Driskell JE. Games, motivation, and learning: a research and practice model. Simul Gaming 2016 Aug 18;33(4):441-467. [doi: 10.1177/1046878102238607]
3. Ruiz-López T, Sen S, Jakobsen E, Tropé A, Castle PE, Hansen BT, et al. FightHPV: design and evaluation of a mobile game to raise awareness about human papillomavirus and nudge people to take action against cervical cancer. JMIR Serious Games 2019 Apr 8;7(2):e8540 [FREE Full text] [doi: 10.2196/games.8540] [Medline: 30958271]
4. Krebs P, Burkhalter J, Fiske J, Snow H, Schofield E, Iocolano M, et al. The QuitIT coping skills game for promoting tobacco cessation among smokers diagnosed with cancer: pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10071 [FREE Full text] [doi: 10.2196/10071] [Medline: 30632971]
5. Santos LH, Okamoto K, Funghetto SS, Cavalli AS, Hiragi S, Yamamoto G, et al. Effects of social interaction mechanics in pervasive games on the physical activity levels of older adults: quasi-experimental study. JMIR Serious Games 2019 Jul 22;7(3):e13962 [FREE Full text] [doi: 10.2196/13962] [Medline: 31333202]

6. Klincova M, Harazim H, Schwarz D, Kosinova M, Smekalova O, Stourac P. What can be achieved with motivation-based teaching of medical students? A monocentric retrospective audit of retention among highly motivated graduates who underwent the learning-by-doing concept in anesthesiology and intensive care medicine. JMIR Serious Games 2019 Apr 9;7(2):e10155 [FREE Full text] [doi: 10.2196/10155] [Medline: 30964443]
7. Schwarz D, Štourač P, Komenda M, Harazim H, Kosinová M, Gregor J, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res 2013 Jul 8;15(7):e135 [FREE Full text] [doi: 10.2196/jmir.2590] [Medline: 23835586]
8. Chon S, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious games in surgical medical education: a virtual emergency department as a tool for teaching clinical reasoning to medical students. JMIR Serious Games 2019 Mar 5;7(1):e13028 [FREE Full text] [doi: 10.2196/13028] [Medline: 30835239]
9. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput Hum Behav 2016 Jan;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
10. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World. New York, USA: Penguin Press; 2011.
11. Annetta LA. The 'I's' have it: a framework for serious educational game design. Rev Gen Psychol 2010 Jun;14(2):105-113. [doi: 10.1037/a0018985]
12. Kapp K. The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. San Francisco, CA: John Wiley & Sons; 2012.
13. Culyba S. Transformational Games: A Process Tool for the Development of Transformational Games. Halifax, NS: ETC Press; 2018.
14. Barab S, Gresalfi M, Dodge T, Ingram-Goble A. Narratizing disciplines and disciplinizing narratives: games as 21st century curriculum. Int J Gaming Comput Mediat Simul 2010;2(1):17-30. [doi: 10.4018/jgcms.2010010102]
15. DeSmet A, van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
16. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
17. Kharrazi H, Lu AS, Gharghabi F, Coleman W. A scoping review of health game research: past, present, and future. Games Health J 2012 Apr 18;1(2):- [FREE Full text] [doi: 10.1089/g4h.2012.0011] [Medline: 24416638]
18. DeSmet A, Shegog R, van Ryckeghem D, Crombez G, de Bourdeaudhuij I. A systematic review and meta-analysis of interventions for sexual health promotion involving serious digital games. Games Health J 2015 Apr;4(2):78-90. [doi: 10.1089/g4h.2014.0110] [Medline: 26181801]
19. Larsen LH, Schou L, Lund HH, Langberg H. The physical effect of exergames in healthy elderly-a systematic review. Games Health J 2013 Aug;2(4):205-212. [doi: 10.1089/g4h.2013.0036] [Medline: 26192224]
20. Proença JP, Quaresma C, Vieira P. Serious games for upper limb rehabilitation: a systematic review. Disabil Rehabil Assist Technol 2018 Jan;13(1):95-100. [doi: 10.1080/17483107.2017.1290702] [Medline: 28359181]
21. Theng Y, Lee JW, Patinadan PV, Foo SS. The use of videogames, gamification, and virtual environments in the self-management of diabetes: a systematic review of evidence. Games Health J 2015 Oct;4(5):352-361. [doi: 10.1089/g4h.2014.0114] [Medline: 26287926]
22. Thomas TH. Protocol and search strategy for a scoping review of serious games for patients with chronic illness. Open Science Framework. 2020 Mar. URL: https://mfr.osf.io/render?url=https://osf.io/k7sy5/?direct%26mode=render%26action=download%26mode=render%20DOI%2010.17605/OSF.IO/48ZHJ [accessed 2020-07-15]
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. About Chronic Diseases. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2020-03-01]
25. Chronic Conditions. Centers for Medicare and Medicaid Services. 2019. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main [accessed 2020-07-08]
26. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Prev Chronic Dis 2013 Apr 25;10:E66 [FREE Full text] [doi: 10.5888/pcd10.120239] [Medline: 23618546]
27. Stokes B. Videogames have changed: time to consider 'serious games'? J Dev Educ 2005;11(3):12 No PMID [FREE Full text]
28. Saposnik G, Levin M, Outcome Research Canada (SORCan) Working Group. Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. Stroke 2011 May;42(5):1380-1386. [doi: 10.1161/STROKEAHA.110.605451] [Medline: 21474804]

29. Burdea G, Polistico K, Krishnamoorthy A, House G, Rethage D, Hundal J, et al. Feasibility study of the BrightBrainer integrative cognitive rehabilitation system for elderly with dementia. Disabil Rehabil Assist Technol 2015;10(5):421-432 [FREE Full text] [doi: 10.3109/17483107.2014.900575] [Medline: 24679074]
30. House G, Burdea G, Polistico K, Grampurohit N, Roll D, Damiani F, et al. A rehabilitation first-tournament between teams of nursing home residents with chronic stroke. Games Health J 2016 Feb;5(1):75-83 [FREE Full text] [doi: 10.1089/g4h.2015.0072] [Medline: 26741697]
31. House G, Burdea G, Polistico K, Roll D, Kim J, Grampurohit N, et al. Integrative rehabilitation of residents chronic post-stroke in skilled nursing facilities: the design and evaluation of the BrightArm duo. Disabil Rehabil Assist Technol 2016 Nov;11(8):683-694. [doi: 10.3109/17483107.2015.1068384] [Medline: 26218427]
32. Kannan L, Vora J, Bhatt T, Hughes SL. Cognitive-motor exergaming for reducing fall risk in people with chronic stroke: a randomized controlled trial. NeuroRehabilitation 2019;44(4):493-510. [doi: 10.3233/NRE-182683] [Medline: 31256084]
33. Simmons CD, Arthanat S, Macri VJ. Pilot study: computer-based virtual anatomical interactivity for rehabilitation of individuals with chronic acquired brain injury. J Rehabil Res Dev 2014;51(3):377-390 [FREE Full text] [doi: 10.1682/JRRD.2013.05.0103] [Medline: 25019661]
34. van Balkom TD, Berendse HW, van der Werf YD, Twisk JW, Zijlstra I, Hagen RH, et al. COGTIPS: a double-blind randomized active controlled trial protocol to study the effect of home-based, online cognitive training on cognition and brain networks in Parkinson's disease. BMC Neurol 2019 Jul 31;19(1):179 [FREE Full text] [doi: 10.1186/s12883-019-1403-6] [Medline: 31366395]
35. Kerfoot BP, Gagnon DR, McMahon GT, Orlander JD, Kurgansky KE, Conlin PR. A team-based online game improves blood glucose control in veterans with type 2 diabetes: a randomized controlled trial. Diabetes Care 2017 Sep;40(9):1218-1225. [doi: 10.2337/dc17-0310] [Medline: 28790131]
36. Gorini A, Marzorati C, Casiraghi M, Spaggiari L, Pravettoni G. A neurofeedback-based intervention to reduce post-operative pain in lung cancer patients: study protocol for a randomized controlled trial. JMIR Res Protoc 2015 May 4;4(2):e52 [FREE Full text] [doi: 10.2196/resprot.4251] [Medline: 25940965]
37. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]
38. Wang H, McMillan SC, Vijayakumar N, McDonald S, Huang L, Gwede C, et al. A behavioral physical activity intervention to manage moderate and severe fatigue among head and neck cancer patients-pre-efficacy study in the national institutes of health orbit model. Cancer Nurs 2019;42(1):E1-14. [doi: 10.1097/NCC.0000000000000568] [Medline: 29461282]
39. Radhakrishnan K, Toprac P, O'Hair M, Bias R, Kim MT, Bradley P, et al. Interactive digital e-health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374 [FREE Full text] [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
40. Fortunato M, Harrison J, Oon AL, Small D, Hilbert V, Rareshide C, et al. Remotely monitored gamification and social incentives to improve glycemic control among adults with uncontrolled type 2 diabetes (iDiabetes): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Nov 20;8(11):e14180 [FREE Full text] [doi: 10.2196/14180] [Medline: 31746765]
41. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675 [FREE Full text] [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
42. Lyons EJ, Baranowski T, Basen-Engquist KM, Lewis ZH, Swartz MC, Jennings K, et al. Testing the effects of narrative and play on physical activity among breast cancer survivors using mobile apps: study protocol for a randomized controlled trial. BMC Cancer 2016 Mar 9;16:202 [FREE Full text] [doi: 10.1186/s12885-016-2244-y] [Medline: 26960972]
43. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. 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 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
44. Höchsmann C, Müller O, Ambühl M, Klenk C, Königstein K, Infanger D, et al. Novel smartphone game improves physical activity behavior in type 2 diabetes. Am J Prev Med 2019 Jul;57(1):41-50. [doi: 10.1016/j.amepre.2019.02.017] [Medline: 31128953]
45. Adie K, Schofield C, Berrow M, Wingham J, Humfryes J, Pritchard C, et al. Does the use of Nintendo Wii Sports improve arm function? Trial of Wii in Stroke: a randomized controlled trial and economics analysis. Clin Rehabil 2017 Feb;31(2):173-185. [doi: 10.1177/0269215516637893] [Medline: 26975313]
46. Adomavičienė A, Daunoravičienė K, Kubilius R, Varžaitytė L, Raistenskis J. Influence of new technologies on post-stroke rehabilitation: a comparison of armeo spring to the kinect system. Medicina (Kaunas) 2019 Apr 9;55(4):- [FREE Full text] [doi: 10.3390/medicina55040098] [Medline: 30970655]
47. Choi JH, Han EY, Kim BR, Kim SM, Im SH, Lee SY, et al. Effectiveness of commercial gaming-based virtual reality movement therapy on functional recovery of upper extremity in subacute stroke patients. Ann Rehabil Med 2014 Aug;38(4):485-493 [FREE Full text] [doi: 10.5535/arm.2014.38.4.485] [Medline: 25229027]

48. Choi Y, Ku J, Lim H, Kim YH, Paik N. Mobile game-based virtual reality rehabilitation program for upper limb dysfunction after ischemic stroke. Restor Neurol Neurosci 2016 May 2;34(3):455-463. [doi: 10.3233/RNN-150626] [Medline: 27163250]
49. da Silva Alves R, Iunes DH, de Carvalho JM, Menezes FD, Silva AM, Borges JB, et al. Effects of exergaming on quality of life in cancer patients. Games Health J 2018 Dec;7(6):385-392. [doi: 10.1089/g4h.2017.0174]
50. da Silva Alves R, Iunes DH, Pereira IC, Borges JB, Nogueira DA, Silva AM, et al. Influence of exergaming on the perception of cancer-related fatigue. Games Health J 2017 Apr;6(2):119-126. [doi: 10.1089/g4h.2016.0051] [Medline: 28418751]
51. da Silva Ribeiro NM, Ferraz DD, Pedreira E, Pinheiro I, da Silva Pinto AC, Neto MG, et al. Virtual rehabilitation via Nintendo Wii and conventional physical therapy effectively treat post-stroke hemiparetic patients. Top Stroke Rehabil 2015 Aug;22(4):299-305. [doi: 10.1179/1074935714Z.0000000017] [Medline: 26258455]
52. Fan S, Su F, Chen S, Hou W, Sun J, Chen K, et al. Improved intrinsic motivation and muscle activation patterns in reaching task using virtual reality training for stroke rehabilitation: a pilot randomized control trial. J Med Biol Eng 2014;34(4):399-407. [doi: 10.5405/jmbe.1502]
53. Gamito P, Oliveira J, Coelho C, Morais D, Lopes P, Pacheco J, et al. Cognitive training on stroke patients via virtual reality-based serious games. Disabil Rehabil 2017 Feb;39(4):385-388. [doi: 10.3109/09638288.2014.934925] [Medline: 25739412]
54. Hickman RL, Clochesy JM, Pinto MD, Burant C, Pignatiello G. Impact of a serious game for health on chronic disease self-management: preliminary efficacy among community dwelling adults with hypertension. J Health Hum Serv Adm 2015;38(2):253-275. [Medline: 26442364]
55. Hung J, Chou C, Chang H, Wu W, Hsieh Y, Chen P, et al. Cognitive effects of weight-shifting controlled exergames in patients with chronic stroke: a pilot randomized comparison trial. Eur J Phys Rehabil Med 2017 Oct;53(5):694-702. [doi: 10.23736/S1973-9087.17.04516-6] [Medline: 28382812]
56. Kempf K, Martin S. Autonomous exercise game use improves metabolic control and quality of life in type 2 diabetes patients - a randomized controlled trial. BMC Endocr Disord 2013 Dec 10;13:57 [FREE Full text] [doi: 10.1186/1472-6823-13-57] [Medline: 24321337]
57. Kim SM, Kim H, Hwang HC, Hong JS, Bae S, Min KJ, et al. The effects of a serious game on depressive symptoms and anxiety in breast cancer patients with depression: a pilot study using functional magnetic resonance imaging. Games Health J 2018 Dec;7(6):409-417. [doi: 10.1089/g4h.2017.0183] [Medline: 30383458]
58. Klompstra L, Jaarsma T, Strömberg A. An in-depth, longitudinal examination of the daily physical activity of a patient with heart failure using a Nintendo Wii at home: a case report. J Rehabil Med 2013 Jun;45(6):599-602 [FREE Full text] [doi: 10.2340/16501977-1151] [Medline: 23695704]
59. Senior H, Henwood T, de Souza D, Mitchell G. Investigating innovative means of prompting activity uptake in older adults with type 2 diabetes: a feasibility study of exergaming. J Sports Med Phys Fitness 2016 Oct;56(10):1221-1225. [Medline: 26329840]
60. Shin J, Bog Park S, Ho Jang S. Effects of game-based virtual reality on health-related quality of life in chronic stroke patients: a randomized, controlled study. Comput Biol Med 2015 Aug;63:92-98. [doi: 10.1016/j.compbiomed.2015.03.011] [Medline: 26046499]
61. Shin J, Kim M, Lee J, Jeon Y, Kim S, Lee S, et al. Effects of virtual reality-based rehabilitation on distal upper extremity function and health-related quality of life: a single-blinded, randomized controlled trial. J Neuroeng Rehabil 2016 Feb 24;13:17 [FREE Full text] [doi: 10.1186/s12984-016-0125-x] [Medline: 26911438]
62. Şimşek TT, Çekok K. The effects of Nintendo Wii(TM)-based balance and upper extremity training on activities of daily living and quality of life in patients with sub-acute stroke: a randomized controlled study. Int J Neurosci 2016 Dec;126(12):1061-1070. [doi: 10.3109/00207454.2015.1115993] [Medline: 26626539]
63. Song GB, Park EC. Effect of virtual reality games on stroke patients' balance, gait, depression, and interpersonal relationships. J Phys Ther Sci 2015 Jul;27(7):2057-2060 [FREE Full text] [doi: 10.1589/jpts.27.2057] [Medline: 26311925]
64. Sutanto Y, Makhabah D, Aphridasari J, Doewes M, Suradi, Ambrosino N. Videogame assisted exercise training in patients with chronic obstructive pulmonary disease: a preliminary study. Pulmonology 2019;25(5):275-282 [FREE Full text] [doi: 10.1016/j.pulmoe.2019.03.007] [Medline: 31076287]
65. Towle D, Edmans J, Lincoln N. Use of computer-presented games with memory-impaired stroke patients. Clin Rehabil 2016 Jul;2(4):303-307. [doi: 10.1177/026921558800200406]
66. van de Weijer SC, Duits AA, Bloem BR, Kessels RP, Jansen JF, Köhler S, et al. The Parkin'Play study: protocol of a phase II randomized controlled trial to assess the effects of a health game on cognition in Parkinson's disease. BMC Neurol 2016 Nov 3;16(1):209 [FREE Full text] [doi: 10.1186/s12883-016-0731-z] [Medline: 27809791]
67. van Santen J, Dröes RM, Bosmans JE, Blanson Henkemans OA, van Bommel S, Hakvoort E, et al. The (cost-) effectiveness of exergaming in people living with dementia and their informal caregivers: protocol for a randomized controlled trial. BMC Geriatr 2019 Feb 19;19(1):50 [FREE Full text] [doi: 10.1186/s12877-019-1062-x] [Medline: 30782127]
68. Yu F, Mathiason MA, Johnson K, Gaugler JE, Klassen D. Memory matters in dementia: efficacy of a mobile reminiscing therapy app. Alzheimer's Dement (N Y) 2019;5:644-651 [FREE Full text] [doi: 10.1016/j.trci.2019.09.002] [Medline: 31720365]

69. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020 Apr 29;8(2):e16096 [FREE Full text] [doi: 10.2196/16096] [Medline: 32347811]
70. Klonoff DC. Behavioral theory: the missing ingredient for digital health tools to change behavior and increase adherence. J Diabetes Sci Technol 2019 Mar;13(2):276-281 [FREE Full text] [doi: 10.1177/1932296818820303] [Medline: 30678472]
71. Rohrbach N, Chicklis E, Levac DE. What is the impact of user affect on motor learning in virtual environments after stroke? A scoping review. J Neuroeng Rehabil 2019 Jun 27;16(1):79 [FREE Full text] [doi: 10.1186/s12984-019-0546-4] [Medline: 31248439]
72. Verschueren S, Buffel C, vander Stichele G. What is the impact of user affect on motor learning in virtual environments after stroke? A scoping review. JMIR Serious Games 2019 May 2;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
73. 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]
74. Kuipers DA, Terlouw G, Wartena BO, van 't Veer JT, Prins JT, Pierie JPE. The role of transfer in designing games and simulations for health: systematic review. JMIR Serious Games 2017 Nov 24;5(4):e23 [FREE Full text] [doi: 10.2196/games.7880] [Medline: 29175812]

Abbreviations

COPD: chronic obstructive pulmonary disease
HbA_{1c}: hemoglobin A_{1c}

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by G Pignatiello, P Stourac; comments to author 12.06.20; revised version received 23.06.20; accepted 25.06.20; published 30.07.20.

Please cite as:
Thomas TH, Sivakumar V, Babichenko D, Grieve VLB, Klem ML
Mapping Behavioral Health Serious Game Interventions for Adults With Chronic Illness: Scoping Review
JMIR Serious Games 2020;8(3):e18687
URL: http://games.jmir.org/2020/3/e18687/
doi:10.2196/18687
PMID:32729836

Original Paper

User Experience of Interactive Technologies for People With Dementia: Comparative Observational Study

Luis Duarte Andrade Ferreira[1,2], MSc; Henrique Ferreira[2,3], MSc; Sofia Cavaco[1], PhD; Mónica Cameirão[1,2,3], PhD; Sergi Bermúdez i Badia[1,2,3], PhD

[1]NOVA LINCS, Dep de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisboa, Portugal

[2]Madeira Interactive Technologies Institute, Universidade Madeira, Funchal, Portugal

[3]Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira, Funchal, Portugal

Corresponding Author:
Luis Duarte Andrade Ferreira, MSc
NOVA LINCS, Dep de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Campus de Caparica, Caparica
Lisboa, 2829-516
Portugal
Phone: 351 21 294 8300
Email: luis.d.ferreira@m-iti.org

Abstract

Background: Serious games (SGs) are used as complementary approaches to stimulate patients with dementia. However, many of the SGs use *out-of-the-shelf* technologies that may not always be suitable for such populations, as they can lead to negative behaviors, such as anxiety, fatigue, and even cybersickness.

Objective: This study aims to evaluate how patients with dementia interact and accept 5 *out-of-the-shelf* technologies while completing 10 virtual reality tasks.

Methods: A total of 12 participants diagnosed with dementia (mean age 75.08 [SD 8.07] years, mean Mini-Mental State Examination score 17.33 [SD 5.79], and mean schooling 5.55 [SD 3.30]) at a health care center in Portugal were invited to participate in this study. A within-subject experimental design was used to allow all participants to interact with all technologies, such as HTC VIVE, head-mounted display (HMD), tablet, mouse, augmented reality (AR), leap motion (LM), and a combination of HMD with LM. Participants' performance was quantified through behavioral and verbal responses, which were captured through video recordings and written notes.

Results: The findings of this study revealed that the user experience using technology was dependent on the patient profile; the patients had a better user experience when they use technologies with direct interaction configuration as opposed to indirect interaction configuration in terms of assistance required (*P*=.01) and comprehension (*P*=.01); the participants did not trigger any emotional responses when using any of the technologies; the participants' performance was task-dependent; the most cost-effective technology was the mouse, whereas the least cost-effective was AR; and all the technologies, except for one (HMD with LM), were not exposed to external hazards.

Conclusions: Most participants were able to perform tasks using *out-of-the-shelf* technologies. However, there is no perfect technology, as they are not explicitly designed to address the needs and skills of people with dementia. Here, we propose a set of guidelines that aim to help health professionals and engineers maximize user experience when using such technologies for the population with dementia.

(JMIR Serious Games 2020;8(3):e17565) doi:10.2196/17565

KEYWORDS

dementia; technology; interaction; psychomotor performance; equipment safety; costs and cost analysis; user-computer interface

Introduction

Background

The difference between serious games (SGs) or transformational games [1] and entertainment video games is that SGs are software apps with a defined goal that goes beyond pure entertainment [2]. In the field of health care, SGs have been developed and used for a variety of purposes, such as training and simulation [3], diagnosis and therapy [4-6], education [7], and other purposes [8]. In recent years, there has been a growing interest in using games to target health conditions, such as stroke [9], Parkinson disease [10], and autism [11]. To a lesser extent, SGs have also been used with people with dementia [12].

Dementia is a neurocognitive disorder [13,14], which impairs cognitive and emotional behaviors, such as memory, language, problem-solving, anxiety, irritability, visuospatial issues, gait and balance-related issues, and other dementia-related aspects [13-18]. Although Alzheimer disease is the most common form of dementia [15], there are other types of dementia, such as vascular dementia [19], Lewy body dementia [20], frontotemporal dementia [21], and mixed dementia [22]. In general, the disease can profoundly affect the carriers, family members, professional caregivers [23], and health care systems [24]. It is estimated that over 35 million individuals worldwide have dementia and that dementia-related expenses reached approximately US $818 billion in 2015 [25]. Portugal alone had expenses ranging between US $1652.8 million and US $2120.4 million in 2009 [24].

Although there are pharmaceutical approaches to treat dementia, these often have side effects, or their desired outcome is only temporary [26]. In addition, the development of new drugs is not only expensive but also time-consuming, as it needs to go through several scientific trials before being approved for human use [27]. As a result, the search for alternative methods, such as SGs, has greatly increased [28].

SGs and Technologies

SGs for people with dementia have been developed as an assistive tool to promote physical, cognitive, and emotional stimulation, leading to a better quality of life [12,29]. Moreover, a novel SG can be used to assess cognitive decline at the early stages (or elevated risk) of dementia [30-32]. Investment in such computer apps can provide an opportunity to reduce institutional health care costs and enhance the quality of life of both family caregivers and people with dementia [33].

Many platform strategies have been considered to develop dementia-related SG apps. Some technologies rely on indirect interactions, such as (PCs) [34], or conventional entertainment systems, such as the Nintendo Wii system [35]. Other technologies are based on direct interaction, such as augmented reality (AR) [36], touchscreen technology [37-39], and gesture recognition systems, such as leap motion (LM), Kinect, and Bracelet Myo [40].

Indirect interaction technologies require an intermediate device to translate human action into interaction with the virtual environment. Indirect interaction devices use more cognitive resources, as they involve conscious spatial and mental translations to convert real-world movements into virtual actions [41].

Using direct interaction technologies, participants do not have an intermediary device to interact with the virtual environment; participants interact directly with the machines using their bodies [41]. In addition, direct interaction devices require less cognitive resources, as there is no movement translation between the real and virtual worlds as opposed to indirect interaction devices [41].

Proposed Frameworks to Develop SGs

In terms of the development of SGs, previous studies revealed interesting insights regarding the development process of SG. For example, Brian Winn proposed the *design, player, experience* framework, which depicts the relationship between the designer and the player's experience [42]. The framework is quite straightforward: the designer designs the game, which is played by the player according to the player's experience. According to this framework, *play* is mediated by experience. Thus, the player's experience (social, cultural, cognitive, and experimental background) influences the design of the game.

The study also considers the learning process in using technological devices to play games, as it can also influence users' game experience. For example, in a recent study, Vallejo et al [43] evaluated the performance of elderly individuals on a set of technologies while performing 2 different tasks. The study concluded that interaction with technology is dependent on the task, the user's experience, and motivation. In addition, the interaction with technology also depends on how intuitive both hardware and software interfaces are for people with dementia [28]; many high-tech technologies can overwhelm people with dementia on a cognitive level, which can affect the learning curve of handling technology [28,44].

Thus, to enhance the user experience for people with dementia while using novel technology, additional guidelines have been suggested to help developers in designing technologies while addressing the need for people with dementia, such as the responding, enabling, augmenting, failure-free (REAFF) framework [45,46]. The REAFF framework focuses on 4 principles: (1) *responding* (technologies should respond to the needs of people with dementia), (2) *enabling* (technologies should improve the quality of life of people with dementia), (3) *augmenting* (technologies should be able to adapt to the reserved skills of people with dementia), and (4) *failure free* (technologies should be as easy to use as possible without discouraging people with dementia).

Another framework—the virtual reality (VR)-Check framework [47]—has been proposed by Krohn et al [47], who evaluated clinical neuropsychology VR apps for cognitive domain specificity (specifically, the cognitive domain being targeted by the VR app), ecological relevance (if the VR app focuses on activities of daily living), technical feasibility (if the VR app is compatible with the desired technologies), user feasibility (if the VR app is feasible to the target population), user motivation (if the VR app engages the users), task adaptability (if the VR app can be adjusted, for instance, in terms of difficulty), performance quantification (if the VR app can objectively quantify the participant's performance), immersive capacities (how immersive is the VR app for participants), training feasibility (if the VR app is suitable to foster cognitive training), and predictable pitfalls (estimating resource-related costs when using the VR app).

Main Purpose of This Study

Despite the existence of several efforts aimed at providing recommendations to develop SGs [48-50], there is still a lack of usability studies that aim to understand how people with dementia interact and accept different types of technologies to perform specific tasks [51,52]. Although elderly individuals are capable of learning and handling new technologies [53], using novel technology can lead to anxious behaviors among elderly populations [54] or lead to undesirable side effects, such as cybersickness [55] and fatigue [28].

To avoid such behaviors, during the prototype playtest phase, it is essential to record the feedback of each player while interacting with the game, as the experience of one player may differ significantly from the experience of another player [42]. In a recent study, Hackner et al [56] analyzed how people with dementia perform different interaction techniques in a tablet, such as a single tap, swipe, and drag-and-drop gestures. The study identified several interaction issues when performing such interaction techniques and presented different solutions to avoid future problems.

Considering the reported potential of SGs as a complementary approach to stimulate people with dementia, the main goal of this study was to better understand how people with dementia accept and interact with *out-of-the-shelf* technologies and how it influences users' game experience while performing different activities. Moreover, this study aimed to find the most suitable technology to design a customizable interactive system that can exploit reminiscence and music therapy in people with dementia. We recruited 12 participants with dementia to perform several activities with different technologies to evaluate their performance while answering 6 research questions (RQs):

- RQ1. Is there a relationship between the patient's profile and user experience?
- RQ2. Is there a relationship between user experience and direct and indirect interaction?
- RQ3. Does any technology elicit more positive or negative emotional responses?
- RQ4. Overall, which technology is better suited for each task?
- RQ5. Which technology is the most cost-effective?
- RQ6. Which technology is less exposed to external hazards?

Following the results of our experiment, we (1) propose a set of guidelines that can help engineers and developers craft better-suited technologies for this population and (2) suggest additional setups of the technologies used to improve user experience in people with dementia.

Methods

Participants

We recruited 12 participants, 3 males and 9 females, with mean age, 75.08 (SD 8.07) years; mean Mini-Mental State Examination score, 17.33 (SD 5.79); and mean schooling, 5.55 (SD 3.30). This was a convenience sample, and the recruitment of the participants was performed by psychologists at the Madeiran delegation of the Portuguese Alzheimer's Association (Table 1). Participants were eligible if they (1) could use upper limbs independently, (2) had an intact hearing, and (3) were in the initial or intermediate stages of dementia. For the last inclusion criteria, we relied on the clinical information available and did not perform any further assessments. The study was approved by the board of the association and followed the standard procedures for research with human participants. Before beginning the experimental trial, all participants (or legal guardians) signed an informed consent form, and permission was granted to film the sessions. After signing the consent form, participants were briefed about (1) the activity objectives and (2) how to handle the technologies. In addition, participants were informed that they could drop out of the experimental trial at any time.

We defined patients' profiles based on their Mini-Mental State Examination (MMSE) [57] scores, age, and years of schooling. The MMSE scores were assessed before the participation of the experimental trial. Only 5 of the participants reported previous experience with technology. For example, participant 1 had experience using a tablet, whereas participants 11 and 12 had experience with PC. Participants 5 and 7 had experience in handling both PC and tablet.

Table 1. Participants' demographics.

Participants	Genders	Age (years)	MMSE[a] score	Schooling (year)	Diagnostics
1	Female	70	25	Fourth	Alzheimer disease
2	Female	85	19	Fourth	Alzheimer disease
3	Female	78	18	Third	Vascular dementia
4	Male	81	17	—[b]	Alzheimer disease
5	Male	67	24	Fifth	Frontotemporal dementia
6	Female	74	12	Third	Alzheimer disease
7	Female	71	14	Fourth	Alzheimer disease
8	Male	82	21	Fourth	Lewy body dementia
9	Female	65	11	Sixth	Alzheimer disease
10	Female	88	10	Twelfth	Alzheimer disease and Parkinson disease
11	Female	77	26	Fourth	Alzheimer disease
12	Female	63	11	Twelfth	Frontotemporal dementia

[a]MMSE: Mini-Mental State Examination.

[b]Participant 4 does not have any formal schooling.

Technologies Used During the Experiment

For each of the following technologies, we selected generic tasks that required different types of interaction, such as (1) manipulating virtual objects, (2) playing musical instruments, (3) moving virtual objects from A to B, and (4) observation of virtual environments.

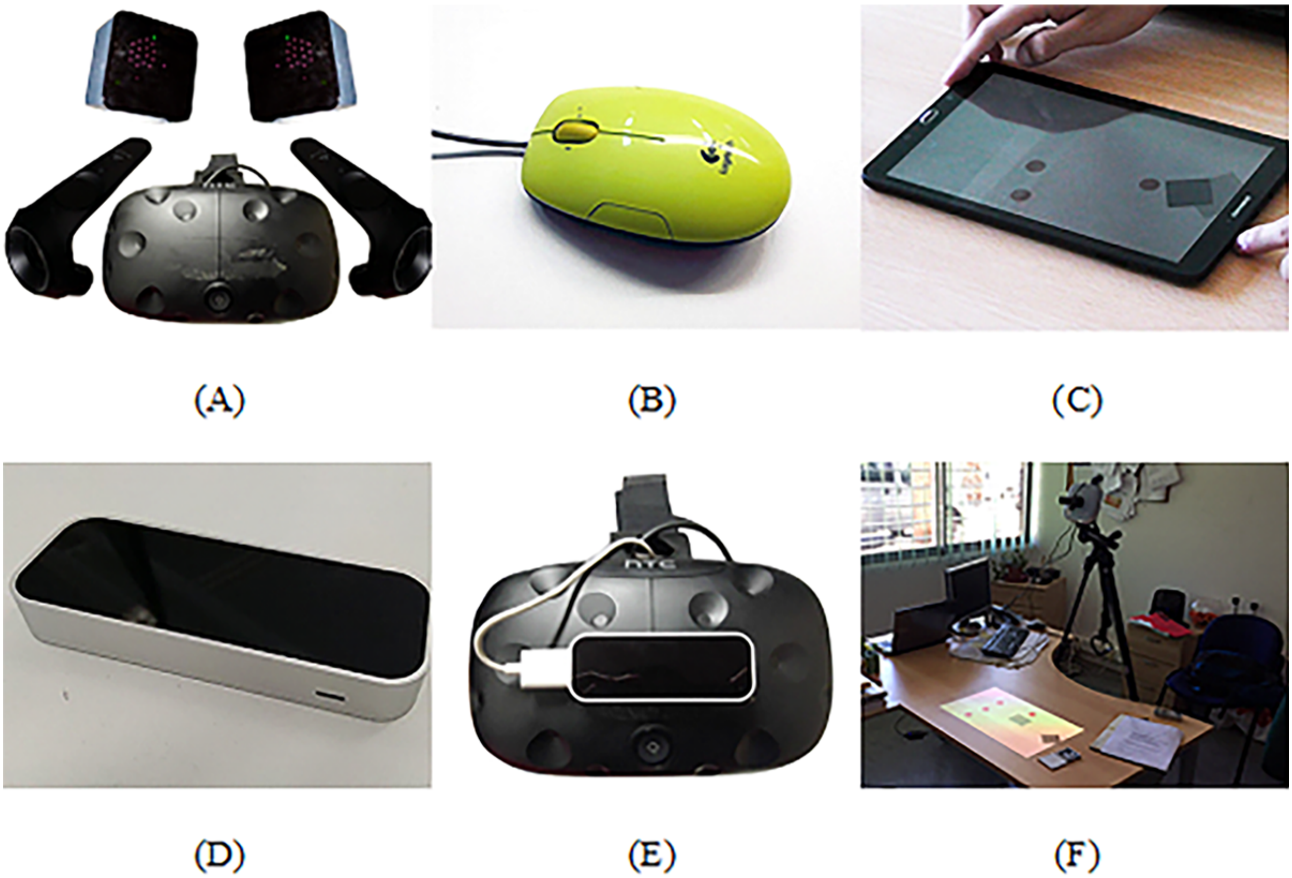
To run the tasks and technologies, we used a Toshiba Satellite L850-1HZ with Windows 10 64 bit equipped with an AMD Radeon HD 7670 and an Intel Core i7-3630QM with 4 GB RAM. Considering that some technologies require a considerable amount of processing power, a desktop computer running Windows 10 64 bits equipped with a Radeon RX 580 Series graphic card and an Intel Core 17-6700 CPU with 16 GB RAM were used. Five different interaction technologies were used in different combinations and tasks.

Indirect Interaction Configurations

Here we present 2 indirect interaction configurations: HMD with controllers and mouse. Each configuration is described below:

1. HTC VIVE with controllers (HMD with controllers): The HTC VIVE technology (HTC) is a set of different technologies that includes a head-mounted display (HMD) and 2 handheld controllers, which are equipped with a trackpad, menu button, system button, trigger button, and grip button. Two base stations were used to track the position and movements of the participant's head and hands (Figure 1).
2. Mouse: We used a standard USB-powered laser mouse (Logitech LS1 Laser Mouse, Logitech International). The mouse is designed with 3 buttons: left, right, and a wheel button (Figure 1).

Figure 1. Technologies used during the experimental trial. (A) HMD with Controllers. (B) Mouse. (C) Tablet. (D) Leap Motion. (E) HMD with Leap Motion. (F) Augmented Reality.



Direct Interaction Configurations

Here we present 5 direct interaction configurations: HTC-VIVE, tablet, LM, HTC with LM and AR. Each configuration is described below:

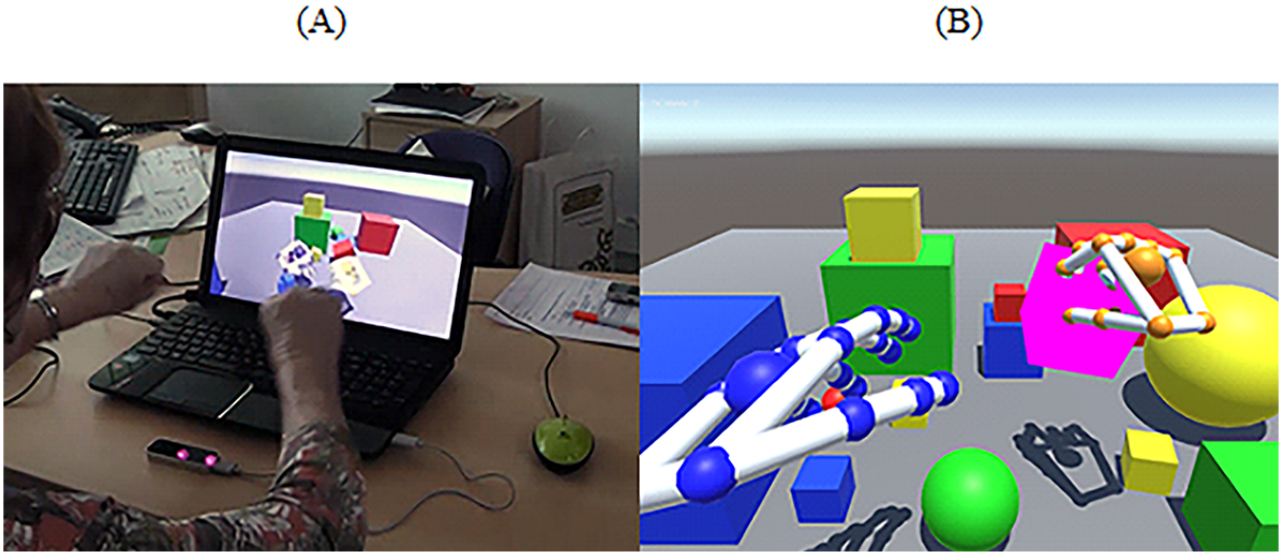
1. HTC VIVE (HMD): The HTC VIVE allows the use of the HMD without the controllers to interact with the virtual environments (Figure 1).
2. Tablet: We used a Samsung 9" Android Tablet (GALAXY, Samsung) that allows interaction inputs, such as tapping and dragging (Figure 1).
3. LM: LM (Motion Control) is an infrared camera–based tracking technology that allows interaction with the virtual environment using hands, fingers, and tools [58] (Figure 1).
4. HTC VIVE with LM (HTC with LM): We added the LM to the HTC VIVE HMD. Thus, participants could interact with the virtual environment not only using head movements but also with their hands (Figure 1).
5. AR: For AR, we developed a projection-based setup that required a projector (LG Inc) and a PlayStation Eye camera (Sony Computer Entertainment Inc), which were attached to a tripod. A physical object with a marker attached to it was used by the participants to interact with the virtual environment. For marker recognition, the Analysis and Tracking System [59] software was used, which allowed the tracking of the physical object (Figure 1).

Manipulating Virtual Objects

LM

The playground was developed using the Unity 3D game engine (Unity Technologies) and consisted of a variety of geometrical figures (Figure 2). In this task, participants were required to use hand gestures, such as grabbing, throwing, and lifting, to interact with the geometrical figures. As participants could interact and throw geometrical figures out of their field-of-view, the virtual playground could be reset by tapping the computer's space bar. The task did not have any music playing in the background, and it did not provide any additional feedback to the participant.

Figure 2. Virtual Playground. (A) Manipulating virtual objects tasks with hands through leap motion. (B) Print screen of the game. Participants had to interact freely with virtual objects.



HMD With Controllers

The goal of the task was to manipulate objects that were placed on a table in a virtual music bar with the HMD with controllers (Figure 3). We used a game called *Jam Session* [60], which can be accessed for free on STEAM (Valve Corporation). Several objects such as cups, a doll, a telephone, a clock, a globe, and a book were used (Figure 3). All objects were placed randomly on the table. To perform the task, participants had (1) to use the controllers with both hands; (2) grab objects by pressing the trigger button on the back of the controller; and (3) rotate, throw, or place the object wherever they wanted. The task did not have any music playing in the background, and it did not provide any additional feedback to the participant.

Figure 3. Steam virtual reality home – bar table. (A) Manipulating daily objects using head-mounted display with controllers. (B) Print screen of the task. Participants had to look at and manipulate daily objects that are on the table.

HMD With LM

For this task, participants had to interact with virtual cubes using both hands while standing (Figure 4). The game is a free demo included in the LM device [61]. The software allows the creation of different kinds of geometrical figures, such as cubes and octagons. Before the beginning of the task, we prepared the scenario by adding multiple geometrical figures in the virtual environment (Figure 4). The goal of the task was to interact with the geometrical figures by making hand gestures, such as grabbing, throwing, or pushing, among other gestures. Participants could interact with either the right or left hand. Participants were positioned in the middle of the room and could move freely around the room. For security reasons, one researcher was always nearby to aid participants whenever needed. No sounds or music was played during the task, and it did not provide any additional feedback to the participant.

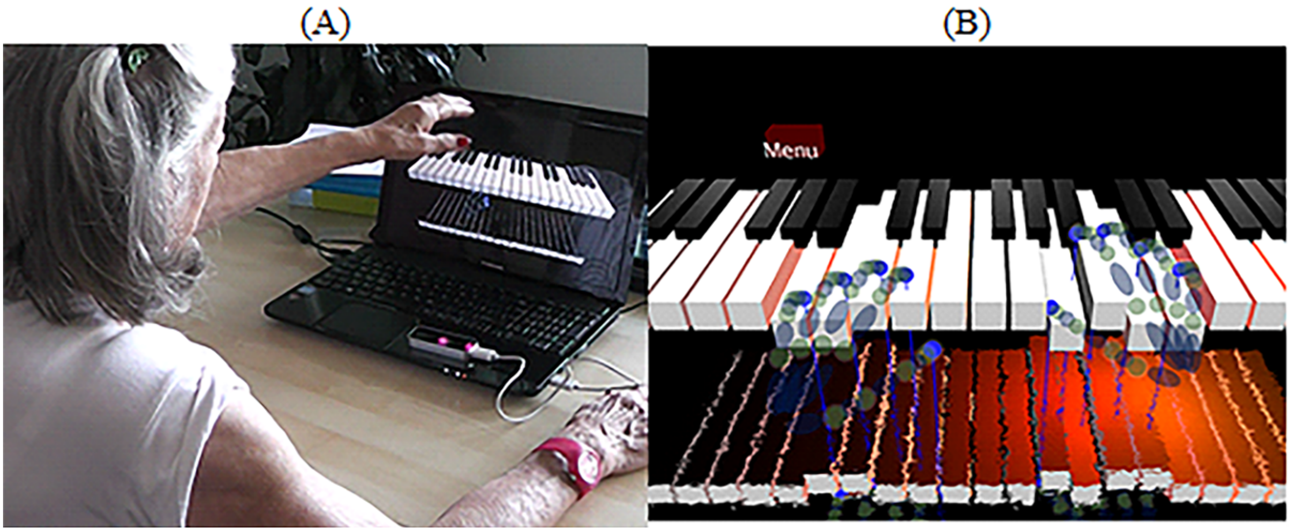
Figure 4. Blocks. (A) Manipulating geometrical figures using hands and head-mounted display with leap motion while standing. (B) Print screen of the game. A set of geometric figures that participants manipulate.

Playing Musical Instruments

LM

The goal of the task was to interact with the piano keyboards using hand movements (Figure 5). The game was a free software included in the LM device [62]. To perform the task, participants had to position their hands above the LM and interact using their fingers. There was no new music playing in the background.

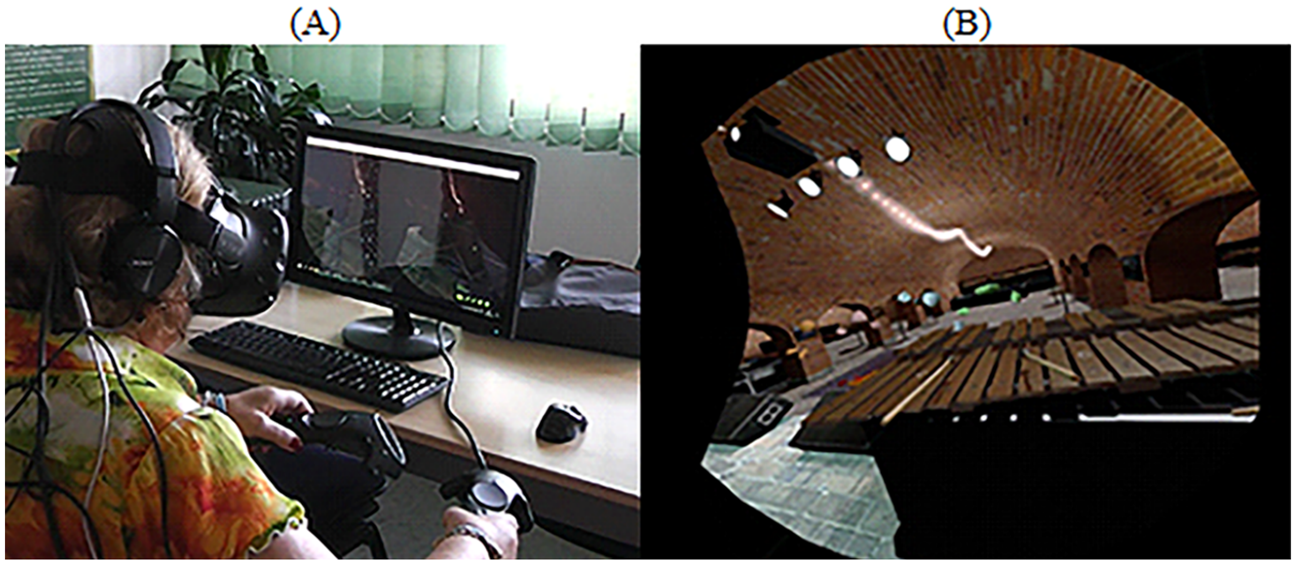
Figure 5. Virtual piano for beginners. (A) Playing the piano using hands and leap motion. (B) Print screen of the game. The virtual piano being played with virtual hands.



HMD With Controllers

We used a virtual environment with a virtual xylophone as, from an interaction perspective, it is very similar to the piano task (Figure 6). We used the free demo of *Jam Session* [60], as it has a variety of instruments, including the xylophone. The goal of the task was to interact with the xylophone while using the HMD headset and handheld controls. To initiate the task, the participant had to (1) grab the controllers with both hands and (2) hit the wooden notes by performing up and down movements with their arms. When interacting with the instrument, dancing avatars would appear in front of the user (Figure 6). Headphones were used by the participants to listen to the sounds while playing the instrument. There was no new music playing in the background.

Figure 6. Steam virtual reality home—Playing musical instruments. (A) Playing the xylophone using head-mounted display with controllers. (B) Print screen of the game. Wooden xylophone with wooden sticks and dancing avatars.

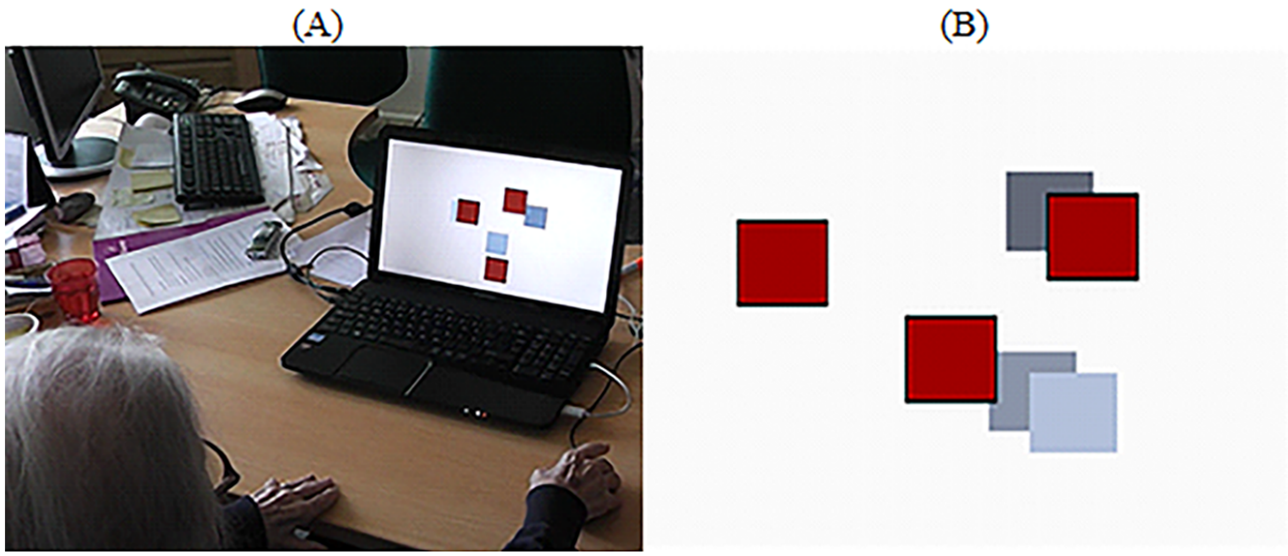


Moving Virtual Objects From A to B

Mouse

The goal of the task consisted of pairing—without any order restriction—a set of 3 randomly placed red squares with 3 randomly placed gray squares using a computer mouse device (Figure 7). The game was custom developed using the Unity 3D game engine (Unity Technologies). To complete the task, the participant had (1) to select a red square by pressing the left mouse button (the square becomes green after selection) and (2) select an available gray square by pressing the left mouse button. The right and wheel buttons were deactivated. Audio feedback was provided with "Very Good!" whenever the participant paired all squares.

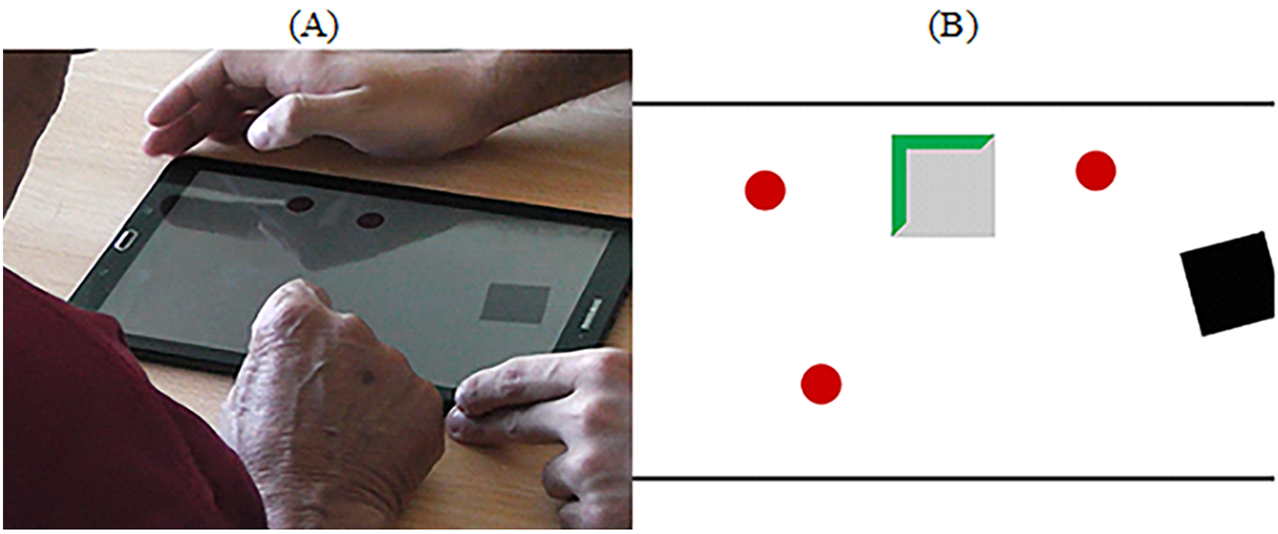
Figure 7. Connecting Squares. (A) Participants are trying to match red squares to gray squares. (B) Print screen of the game. A set of randomly distributed squares.



Tablet

In contrast to the previous task using the mouse, the goal of this task was to capture a set of randomly placed red spheres (Figure 8). The game was developed using the Unity 3D game engine (Unity Technologies). To complete the task, participants had to (1) drag a gray container to a red sphere, (2) wait for 4 seconds to attach the sphere to the container, and (3) drag the container with the sphere attached to it to a black rotating target (Figure 8). A countdown sound would provide feedback during the 4-second countdown. After that, the red sphere would become green. In addition, the participant was rewarded with audio feedback—"Very Good"—when all spheres were captured.

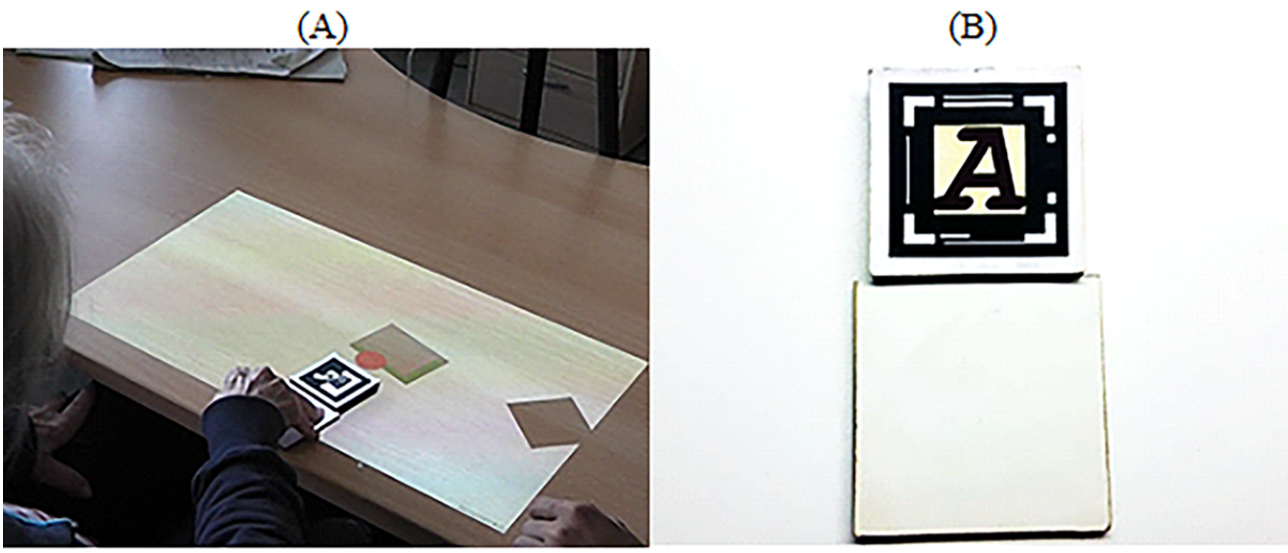
Figure 8. Dragging spheres. (A) Participants are collecting red spheres using a gray container and dragging these with the finger to a black rotating square. (B) Print screen of the game. A set of randomly distributed red spheres, a gray container with the activated timer (in green), and a black rotating target.



AR

For this technology, we used the same task as the one developed for the tablet. However, in this case, participants had to drag a physical object with a gray virtual container attached to it to collect the red spheres and bring them to the black rotating target (Figure 9).

Figure 9. Dragging spheres. (A) Participants are collecting red spheres using a physical object. (B) Physical objects that participants use to interact with the spheres.



Observation

HMD

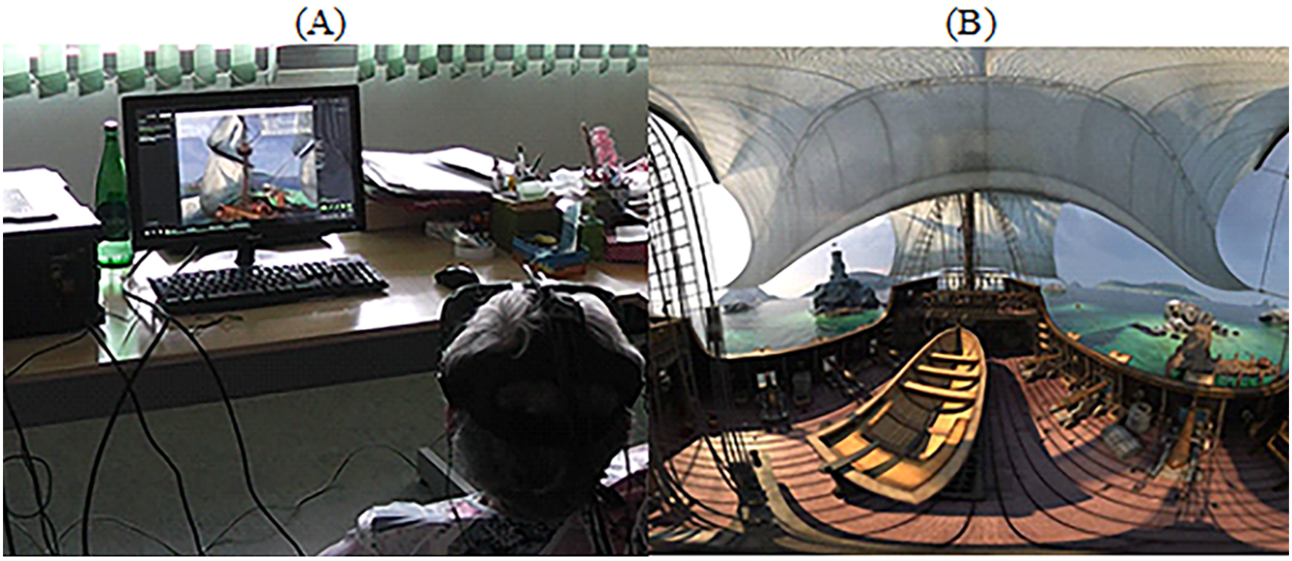
In this case, the participants were seated on a chair while wearing the HMD device. The participants were invited to explore virtual worlds by freely moving their heads. We used 2 different tasks to evaluate the participants' performance:

1. Static scenario—Exploring the forest: In this task, we used a virtual forest that was developed in the Unity 3D game engine (Unity Technologies; Figure 10). It has virtual elements, such as trees, grass, and clouds (among other elements), as well as audible elements, such as birds, insects, and wind (Figure 10). The windy sound effect, combined with the animation of virtual elements, offered dynamism to the scenario by providing the illusion that the virtual elements were moving because of the wind. The goal of this task was to report and describe as many elements as possible.
2. Dynamic scenario—Exploring the ghost ship: This task was a short virtual video of a pirate ship navigating in the natural elements such as rocks, small buildings, and highly detailed pirate ships (Figure 11). The game can be accessed for free on STEAM [63].

   When the video begins, a virtual camera automatically moves on a predefined path while rotating on its axis to show places of interest to the viewer. Generic background music plays in the background, accompanying the participants' journey. Participants were free to move their head to explore the virtual environment. The goal of the task was to describe and report as many virtual elements as possible to the researchers.

Figure 10. Observing virtual environments: Virtual Forrest. (A) Participants observe a virtual forest using the head-mounted display. (B) Screenshot of the game displaying a virtual forest with trees, grass, and sound.

Figure 11. Observing the virtual environments: Ghost Ship. (A) Participants observe an interactive video using the head-mounted display. (B) Screenshot of the Ghost ship sailing in the Caribbean Sea.



Procedure

A within-subject experimental design was used to allow all participants to interact with all technologies and tasks. Each week, different technologies and tasks were randomly introduced. Participants were required to complete tasks, such as manipulating virtual objects, moving virtual objects from A to B, observation of virtual scenarios, and playing musical instruments. Participants were seated in a quiet room and accompanied by 2 researchers and a health professional when needed.

During the experiment, patients sat in front of a table in a silent room of the Madeiran delegation of the Portuguese Alzheimer's Association (except when performing the task requiring a HMD with LM, which in this case required standing up). Two researchers were present in the room; one researcher was responsible for filming and taking notes on the participants' performance, whereas the other researcher interacted with the participants during the experimental trial. The video recordings allowed us to analyze participants' behaviors and study their verbal responses throughout the experiment. The camera was placed behind the shoulders of the participants to conceal their faces and protect their identity. In the case of participants in more intermediate stages of dementia, a health professional was also present to guarantee their well-being and aid researchers during the intervention.

During the experiment, 3 protocols were used: (1) before initiating the task, the researchers instructed participants on how to use a specific technology to complete a task; (2) each task had a maximum duration of 15 min, and participants could repeat tasks if desired; and (3) during the task, participants were encouraged to think aloud and could ask for help at any time. All interventions by the researchers were annotated for later analysis.

Analysis

To address our RQs, we relied on direct observations and behavioral and verbal responses extracted from the video recordings. To analyze the video recordings, we used Adobe Premiere CC 2017.1.2 release version for coding. This video editing tool allowed us to tag, comment, and export annotations in comma-separated values (.CSV) files. The video analysis went through 2 phases. In the first phase, 2 researchers performed independent video analyses by tagging and annotating events in the video files. In the second phase, the information gathered by both researchers was compared and checked for consistency. In case of disagreement, a third researcher was invited to disambiguate.

To analyze participants' user experience with a given technology or in each task, we counted the number of issues identified. The issues were grouped into (1) assistance provided by researchers, (2) perception issues, (3) comprehension issues, (4) interaction issues, and (5) discomfort that participants felt. These are described in detail below.

1. Assistance provided: We counted the number of times participants required assistance from the researchers. In addition, we considered the assistance provided by therapists if they were present during the experimental session.
2. Comprehension issues: We counted all issues identified in terms of the participants' general understanding of how to perform the tasks.
3. Perception issues: We considered (1) visual perception issues whenever participants had difficulties in visualizing and correctly identifying game elements during user experience and (2) sensory issues whenever participants had difficulties in hearing and identifying sounds correctly. For (3) tactile issues, we counted the number of times that participants complained of not feeling any physical feedback of the technology during the user experience (ie, lack of vibration and not finding the correct button) and number of times participants were expected to interact with the virtual environment in the same way as in a real-life scenario (ie, expecting to be able to physically touch and feel a virtual object when interacting with it).
4. Interaction issues: We considered issues such as (1) controlling the interface (ie, clicking incorrect buttons to fulfill a task), (2) controlling the software (ie, triggering wrong software functionalities), and (3) the participants physically misusing the interface (ie, grabbing the LM).
5. Discomfort: We counted the number of times participants felt distressed (ie, fatigue, cybersickness, and balance issues).

In addition, we studied emotional responses (positive and negative). That is, we counted the number of positive emotions (ie, laughter) and negative emotions (ie, frustration). We also counted the number of software issues (ie, undesirable features or *bugs*) that occurred during the experiment. For RQ5, we excluded software issues, as these were explicitly related to the actual software and not to the technology per se.

Finally, we calculated the number of times the equipment was exposed to external hazards—equipment at risk. For example, we counted the number of times the equipment was at risk of falling to the ground during user experience.

Data Analysis

Data were analyzed using SPSS, version 24 (IBM Corp). For each dependent variable, the normality of the distribution was assessed using the one-sample Kolmogorov-Smirnov test. As most distributions deviated from normality, nonparametric statistical tests were used for the analysis. Descriptive results are presented as median and IQR. For assessing the impact of experimental conditions, the Friedman test was used. For post hoc pairwise comparisons, the Wilcoxon signed-rank test was used. The significance level was set at α=.05. Bonferroni correction was used to account for multiple comparisons. We also used Bonferroni correction for analyzing which combinations of technologies and tasks minimized feasibility performance (RQ4) and for analyzing which technology was exposed less to external hazards. For nonparametric correlations, we used Spearman rank correlation coefficient.

Results

Participants

Participants 1, 2, 3, 7, 11, and 12 completed all 10 experimental conditions. Participants 9 and 10 withdrew from the experiment, and participants 4, 5, 6, and 8 were not able to complete all tasks. Consequently, only 9 datasets were considered for the playground activity with LM; 10 datasets were considered for condition LM (piano activity), tablet, and PC. For AR and observation (exploring the forest), 11 datasets were considered, whereas 7 were considered for condition observation (exploring the ghost ship), playing musical instruments, and manipulating virtual objects with both HTC with controllers and HTC with LM, respectively. In addition, some video recordings were corrupted, which did not allow us to computerize the number of issues; instead, we relied on the written notes taken during the experimental trial.

Analysis

RQ1: Is There a Relationship Between the Patient's Profile and User Experience?

We studied the relationship between the patients' profiles when considering each performance domain and each technology separately (Multimedia Appendix 1).

A positive correlation between patients' MMSE score and perception-related issues when using LM (r_s=.652; n=10; *P*=.04) was found. We also found a significant and negative correlation between participants' MMSE scores and the number of assistances provided (r_s=−.744; n=11; *P*=.01) when using the AR technology. In addition, participants' years of schooling correlated negatively (r_s=−.615; n=11; *P*=.04) with perception issues in the AR setup. In terms of the tablet, we found a significant negative correlation in the MMSE scores with both comprehension (r_s=−.726; n=10; *P*=.02) and interaction issues (r_s=−.642; n=10; *P*=.045). Finally, for the HMD with LM, we identified a negative correlation between the MMSE scores and the number of assistances provided (r_s=−.802; n=7; *P*=.03). Multimedia Appendix 2 shows the correlation plots for some of the stronger associations. However, the significance mentioned previously does not endure if adjusted for multiple testing using Bonferroni correction.

Finally, to understand the relationship between the patient profile and performance, we ran a Spearman correlation analysis considering patients' profile—MMSE, age, schooling, and the total number of issues identified (during user experience)—assistance provided, comprehension issues, interaction issues, perception issues, and discomfort. This analysis did not identify statistically significant correlations between user experience and patient profile (Multimedia Appendix 3).

To identify whether there are any specific tasks or technologies where the cognitive profile may play a role, we repeated the analysis on each task (playing musical instruments, manipulating virtual objects, move objects from A to B, and observation) and each type of technology (LM, HMD, AR, tablet, PC, HMD with controllers, and HMD with LM). Again, we did not find any significant correlations for either task type or technology (Multimedia Appendices 4 and 5).

When considering performance scores by the individual performance domains (ie, assistance provided, discomfort as well as comprehension, interaction, and perception issues), we also found no direct association with the patient's cognitive profile (Multimedia Appendix 6).

RQ2: Is There a Relationship Between User Experience and Direct and Indirect Interaction?

In RQ2, we examined whether there was a difference in participants' user experience while using direct (LM, AR, tablet, HMD with LM, and HMD) or indirect interaction technologies (HMD with controllers and mouse). In general, participants required less assistance and were able to understand better how to use direct interaction technologies. More concretely, participants required significantly more assistance using indirect interaction devices (median 3.00, IQR 12.00) than using direct interaction devices (median 1.70, IQR 7.00; Z=−2.666; *P*=.01; r=−0.6). Moreover, participants had significantly more comprehension issues with indirect interaction (median 4.00, IQR 5.50) than with the direct interaction (median 2.00, IQR 2.77; Z=−2.601; r=−0.6, *P*=.01). No statistically significant differences were found in interaction issues (median 8.50), perception issues (median 1.50), and discomfort (median 0.40).

RQ3: Does Any Technology Elicit More Positive or Negative Emotional Responses?

We evaluated participants' overall emotional responses while using each technology. For this analysis, we considered the number of positive minus the number of negative emotional reactions identified in the video analysis. There were no statistical differences between the emotional responses and technologies used (χ^2_6=7.1; *P*=.31).

RQ4: Overall, Which Technology is Better Suited for Each Task?

We analyzed which combinations of technologies and tasks minimized the identified performance and maximized positive emotional reactions. When tasks were grouped by the technology used, participants' comprehension (χ^2_6=23.1; *P*=.001), interaction (χ^2_6=19.6; *P*=.003), and discomfort (χ^2_6=22.9; *P*=.001) were significantly impacted by technology but not by the number of assistances (median 2.00) and perception issues (median 1.00). A post hoc analysis did not reveal any significant pairwise differences. Table 2 shows the ranking of technology in terms of issues. We ranked the technologies according to their median (IQR).

Table 2. Ranking of technologies according to performance domains.

Ranks	Comprehension issues	Interaction issues	Discomfort
First	HMD[a]: 0.00 (0.00)	HMD: 0.00 (0.00)	• Mouse: 0.00 (0.00) • AR[b]: 0.00 (0.00)
Second	HMD with LM[c]: 0.00 (1.00)	AR: 2.00 (7.00)	• N/A[d]
Third	HMD with controllers: 1.00 (5.00)	HMD with LM: 4.00 (6.00)	• Tablet: 0.00 (.50)
Fourth	LM: 1.50 (2.50)	HMD with controllers: 6.00 (11.00)	• HMD: 0.00 (1.00)
Fifth	AR: 2.00 (3.00)	Mouse: 8.50 (14.50)	• HMD with controllers: 0.00 (4.00)
Sixth	Tablet: 4.50 (6.50)	LM: 9.50 (8.25)	• LM: 1.00 (3.25)
Seventh	Mouse: 5.00 (5.25)	Tablet: 24.00 (23.00)	• HMD with LM: 1.00 (5.00)

[a]HMD: head-mounted display.

[b]AR: augmented reality.

[c]LM: leap motion.

[d]N/A: not applicable. Following a standard competition ranking, there is no device ranking second.

Playing Musical Instruments

For this task, we used LM and HMD with controllers, and participants played 2 virtual musical instruments: a piano and a xylophone. Participants showed more perception issues while using the HMD with controllers (median 1.00, IQR 9.00) than when using LM (median 0.00, IQR 0.00; Z=−2.226; r=−0.6, *P*=.03). No other differences between technologies were found.

Manipulating Virtual Objects

For this task, participants used the LM, HMD with LM, and HMD with controllers to manipulate a variety of virtual objects. Participants' performance differed significantly in terms of software issues (χ^2_2=6.3; *P*=.04) and equipment at risk (χ^2_2=6.5; *P*=.04). We did not find differences in terms of assistance (median 1.00), emotional responses (median −1.00), comprehension (median 0.00), perception (median 0.00), and interaction (median 5.00) issues as well as discomfort (median 1.00). Post hoc analysis revealed no significant pairwise differences among conditions. Table 3 shows the ranking of technologies in the domains in which significant differences were identified. Overall, the combination of HMD with controllers shows a more stable performance in this task. We ranked the technologies according to their median (IQR).

Table 3. Ranking of participants' performance to manipulate objects.

Ranks	Software issues	Equipment at risk
First	HMD[a] with controllers: 0.00 (0.00)	• LM[b]: 0.00 (0.00) • HMD with controllers: 0.00 (0.00)
Second	HMD with LM: 1.00 (3.00)	• N/A[c]
Third	LM: 1.00 (5.00)	• HMD with LM: 1.00 (1.00)

[a]HMD: head-mounted display.

[b]LM: leap motion.

[c]N/A: not applicable. Following a standard competition ranking, there is no device ranking second.

Moving Virtual Objects From A to B

For this task, participants used tablet, AR, and mouse devices to move objects from A to B. We found a significant effect of technology in software issues (χ^2_2=13.0; *P*=.002) but not in assistance (median 2.00), emotional responses (median 1.00), comprehension issues (median 4.00), interaction issues (median 8.00), perception issues (median 2.00), and discomfort (median 0.00). The technology that raised more software issues was AR (median 2.00, IQR 3.00), followed by tablet (median 0.50, IQR 1.25) and mouse (median 0.00, IQR 0.00). However, no significant pairwise differences were found among them.

Observation

In this task, we studied the impact of 2 modalities: static versus moving content on HMD. Participants explored 2 different environments: a virtual forest and an interactive video. No differences were identified between the 2 modalities. Figure 12 summarizes the findings, reporting the most appropriate technologies by task—manipulating virtual objects, moving virtual objects from A to B, playing musical instruments, and observation.

RQ5: Which Technology is the Most Cost-Effective?

One critical factor that may limit the adoption of interactive technologies in this area is their cost. Hence, it is essential to perform a cost-effectiveness analysis to inform therapists and caregivers on the implications of their technological choices in terms of costs and outcomes. In this study, the most expensive technologies were HMD with LM (€578.99 [US $661.76]) and AR (€523.54 [US $598.38]), whereas the cheapest ones were the mouse (€16.99 [US $19.42]), LM (€79.99 [US $91.42]), and tablet (€79.99 [US $91.42]). HMD (€499.00 [US $570.33]), and HMD with controllers (€499.00 [US $570.33]) technology presented a moderate cost. In terms of the (accumulated) identified issues during the study, HMD (46 issues) and HMD with LM (51 issues) had the least issues, whereas tablet presented the most performance issues (433 issues). Technologies such as mouse (209 issues), HMD with controllers (158 issues), LM (166 issues), and AR (193 issues) presented intermediate performance issues. A cost-effectiveness analysis aims to find the right balance that minimizes both cost and number of issues (Multimedia Appendix 7).

We multiplied the number of issues with the purchase price of each technology to calculate the cost efficiency of each technology. The results are presented as the absolute value between the identified issues and costs. As we can see in Multimedia Appendix 7, the most cost-efficient technology is the mouse device, whereas AR is the least cost-efficient technology.

Figure 12. Suitable technologies for each task. Grayscale intensity represents the total number of issues (the lower the intensity, the lower the number of errors), and x represents technologies that have not been used to perform that given task. LM: leap motion; HMD: head-mounted display; AR: augmented reality.

Task	Mouse	Tablet	LM	HMD	HMD with LM	AR	HMD with controllers
Manipulating virtual objects	×	×		×		×	
Playing musical instruments	×	×		×	×	×	
Moving virtual objects from A to B			×	×	×		×
Observation static	×	×	×		×	×	×
Observation interactive video	×	×	×		×	×	×

RQ6: Which Technology is Less Exposed to External Hazards?

With all the technologies used in this study, we analyzed how they were exposed to risks that could damage the equipment. We found a statistical, but very modest, effect of the type of technology (χ^2_6=15.9; *P*=.01). The technology that led to higher risk situations was the HMD with LM (median 1.00, IQR 1.00). However, post hoc pairwise comparisons revealed no significant differences among technologies.

Discussion

Comprehension

The technology that ranked best in terms of comprehension was the HMD, whereas technologies that scored worse were mouse and tablet. This is probably because of the simplicity of the interaction with HMDs—participants only need to move their heads to interact with the virtual environments. However, when using the mouse, participants showed great difficulties in understanding how to use it. Most of the difficulties were related to the mapping of the mouse, and sometimes, participants lost sight of the mouse cursor. Participants also had difficulties in interacting with the buttons, being distracted by the mouse wheel many times, as it is the most salient button of the device. Participants tended to rotate and click it instead of using actual mouse buttons. Some participants tried to rotate the mouse wheel forward and backward to move the mouse cursor up and down on the screen. Such behaviors even occurred in participants who had previous professional experience with it. For example, participant 12 had previous experience using the mouse, yet was unsuccessful. As a result, the participant cried, and the experiment had to be stopped. Thus, it becomes crucial to develop intuitive interfaces to avoid overwhelming participants in understanding how to use technology to complete virtual tasks [28].

Interaction

In terms of interaction, the HMD again ranked the highest. Participants presented issues with the tablet's interface, mostly because of the multi-touch control. When using multi-touch control, participants would tend to rest their hands on the tablet surface and trigger undesired functions that would prevent them from achieving their goal. Once more, an intuitive software interface is vital to enhance performance in people with dementia [28]. As our tablet was not fixed to the table, it also moved around as participants interacted with it. A better setup would have the tablet fixed to a surface, as in the study by Hackner and Lankes [56]. Despite these issues, the participants were able to perform the task gracefully.

Discomfort

Concerning discomfort, participants complained the most when using the LM and HMD with LM. For example, participants 1, 2, 10, and 11 reported fatigue while using LM. Indeed, to interact with the LM, participants' arms need to be moving in the air, leading to muscle fatigue. In the case of the HMD with LM, only participant 5 did not report any discomfort. The remaining participants reported fatigue, cybersickness, and balance difficulties. Although the HMD alone did not trigger major issues, participants 6 and 3 felt nauseated, and participant 12 reported cybersickness after the virtual video task. Participant 6 complained about the heat generated by the headset. In general, cybersickness and fatigue are some of the negative aspects identified in the scientific literature in terms of the use of technology, whereas balance-related issues are associated with the negative consequences of dementia [15,17,28,55].

Effect of Patient Profile

We found that the participants' profile influences the usage of technology. A negative and significant correlation between MMSE scores and the number of assistances provided with AR and HMD with LM were found. In the case of AR technology, we found a significant effect on the level of schooling and the number of perception issues that arose in the experiment. We also saw that a low level of schooling and lack of experience with novel technology could lead to confusion (or even anxiety) [54]. For example, participant 3 was confused when instructed to move the red spheres that were projected on the table; as a result, the participant questioned: "*How can I catch the spheres if they are fixed on a table?*" Concerning the usage of the tablet, we found a significant correlation between MMSE scores and both comprehension and interaction issues. This is likely because of the multi-touch feature. Some participants failed to understand that by placing the whole hand on the screen of the tablet, multi-touch is triggered. Other issues that were identified included (1) activating the menu buttons of the tablet involuntarily, (2) dragging the tablet involuntarily while interacting with the virtual objects, (3) forgetting to wait for the selection time, and (4) forgetting the task rules. Finally, we found a positive and significant correlation between the MMSE scores and the number of perception issues when using LM. We observed that participants with high MMSE scores were able to interact with technology easily and for longer, which allowed researchers to identify perception issues during user experience, in contrast to participants with lower MMSE scores who struggled to begin a given task. Similar results were found in the study by Alvseike and Brønnick [44], which found that individuals with higher cognitive deficits had more difficulties in using smart house technology than individuals with lower cognitive deficits. Performance may also depend on other variables, such as motivation and experience [43].

Direct Versus Indirect Interaction

Participants required more assistance statistically and had more difficulties in understanding how to use indirect interaction devices. Indirect interaction devices require more cognitive resources [41] and, in a population with cognitive deficits, may hinder performance during the completion of tasks. Conversely, direct interaction devices require less cognitive resources, and, consistent with our observations, participants had fewer complications in using such technologies as they are more intuitive and straightforward to interact with virtual content. Some participants, such as participants 1 and 11, were able to use both direct and indirect interaction technologies with minor problems. However, it is important to take into consideration that these participants had higher MMSE scores, and that participant 11 had experience in using mouse technology.

Emotional Responses

Participants, in general, did not show many emotional responses when using the studied technologies. However, some interesting reactions were observed. For example, participant 1 was very happy when she was able to grab a cube while using the HMD with LM and said, "*Oh good...what a funny thing...it is so beautiful.*" The same participant showed pride while playing the xylophone with the HMD with controllers and said that it was a shame that the people in the room could not hear her playing as it was a beautiful song. Participant 11 enjoyed exploring virtual environments with the HMD. She repeatedly said "*very beautiful*" in both the *Exploring the Forest* and *Exploring the Ghost Ship* tasks.

Playing Musical Instruments

Here, participants used the LM and HMD with controllers to play virtual instruments and showed more perception issues while using the HMD with controllers than LM. Most of the issues identified were visual, auditory, and tactile related. For example, participant 12 complained that she did not hear the xylophones (yet, she confirmed during the experience that she heard the sounds). The same participant also reported that she was not able to see anything several times. In addition, participant 3 complained that she was not able to see or reach the musical instruments (despite being within the participant's arm range).

Manipulating Virtual Objects

In this task, the participants used the LM, HMD with LM, and HMD with controllers to manipulate virtual objects. We found differences in terms of software issues and equipment at risk. In general, the best technology is HMD with controllers. Although there were no statistically significant perception issues, participant 12 raised most visual-related problems, as she had difficulties in identifying the virtual objects in the virtual environments, including the digital representation of her hand. Participant 3 complained because she was expecting to "physically" grab the virtual objects. In terms of software issues, the HMD with controllers scored first place as it did present minor issues.

In contrast, LM technology scored the worst (last place). As participants tried to grab virtual objects, sometimes the objects stayed attached involuntarily to their hands, and they struggled to let go of the objects. Similar behaviors were recorded while participants performed the task while using the HMD with LM. Participants were able to *grab* virtual objects but had more difficulties dropping them. Finally, in terms of equipment at-risk situations, the HMD with LM triggered more dangerous situations for the equipment. For example, when participants were performing abrupt movements with the head, the HMD was sometimes at risk of falling.

Move Objects From A to B

We only found a statistical difference in terms of software issues, with AR and tablet being the ones that scored the worst. AR technology had some camera tracking issues because of environmental issues, such as shadows and reflections. In contrast, most of the issues related to the tablet were because of software bugs. Despite these minor issues, all technologies performed at an acceptable level.

Observation

In this task, we did not find any statistical differences. The only issues identified were related to cybersickness in both observation tasks [55].

Design Recommendations

In this study, we observed that technology had different outcomes in terms of acceptance and performance on people with dementia. Although technologies have been accepted by the majority, some participants had difficulties in managing them to fulfill the tasks. Such differences in the results are mainly because of patient profiles, which, in turn, influence technology configuration (direct interaction versus indirect interaction).

Comprehensibly, most of the technologies used were not aligned with the REAFF framework, as these were not explicitly designed to take into consideration the needs of people with dementia [45,46]. Most of the technologies used did not follow, for example, the augmenting or failure free principles, as participants did not complete the tasks independently. It is also important to consider how to align such technologies with the remaining principles of the REAFF framework for the needs of people with dementia (responding) and how technology can improve their everyday life (enabling).

In addition, it would be interesting to re-evaluate such technologies using similar tasks as presented in this study, but in a clinical context using the VR-Check framework [47]. Thus, more detailed knowledge could be gained regarding the adequacy and therapeutic outcome when using technology and virtual reality with people with dementia.

Although the technologies used are not perfectly aligned with the REAFF framework principles, they are accessible and can be used in their favor if they are set up correctly. By studying the use of the different technologies and tasks by people with dementia, we can provide a set of recommendations for the selection and implementation of different technological solutions when working with this population. Table 4 addresses the main problems encountered and provides recommendations to overcome them. These recommendations can help engineers in the design of technologies for people with dementia and draw attention to health professionals and informal caregivers regarding potential issues that can emerge while using such technologies with this population.

Table 4. Identification of problems and proposed recommendations or using technology to perform virtual tasks by people with dementia.

Technology, Identified problems	Solution
LM[a]	
Grabbing/moving technology needlessly	Design a setup where the LM is fixed and not graspable (ie, a 3D printed container or embedded onto the tabletop surface).
Confusing virtual objects (spheres) with the joints of the virtual hand	Use identifiable virtual objects and representations of the hand with higher realism.
Tablet	
Moving the whole tablet involuntarily	Secure the tablet on a table or fixed structure such that patients do not need to hold it and can interact with its touch screen.
Triggering undesired touch inputs	Deactivate multi-touch and disable system buttons.
AR[b]	
Interaction with physical elements	Ergonomic design with affordances consistent with the task at hand can enhance performance.
Tracking problems	The most common tracking problems are related to (1) shadows, (2) markers out of the camera's field-of-view, or (3) projection of virtual elements on markers. Solutions include using a room without direct sunlight and controlled light conditions; using lower contrast virtual elements that diminish interference of projecting on markers; and using a setup with clearly defined interaction boundaries.
Mouse	
Buttons not salient	Select a computer mouse that visually clearly identifies where those buttons are. A large colored sticker or paint on a button can also be used to improve its saliency.
Too many buttons	Most modern computer mice consist of 3 buttons and a scroll wheel. Choose a one-button mouse (ie, Apple mouse). Disabling or mapping all mouse buttons to the same functionality will minimize the impact of choosing the wrong button.
Mouse cursor (and other elements) too small	Increase the size of the mouse cursor and other virtual elements to enhance performance.
HMD[c] with controllers	
Too many buttons in handheld controls	Users only see a virtual representation of the controls in the HMD. Minimum button input should be considered while the remaining buttons are disabled or mapped to the same function.
Hitting controls against each other	Use only one control to interact with the virtual content when possible. Alternatively, replace the controllers with an LM.
HMD with LM	
Lack of haptic feedback	Complement with alternative channels to convey haptic feedback (ie, auditory or visual).
Cybersickness and balance issues	People with disabilities need to be assessed for balance, and seating setups should be considered. Safety harnesses or other safety measures should be considered when standing.
HMD	
Discomfort because of the device's heat	Use in a properly ventilated room. In case of discomfort, divide the session into multiple shorter intervals.
Cybersickness	Virtual environments should be designed to minimize optic flow, and incongruency between physical and virtual motion should be minimized. It can be achieved by reducing forward motion and rotations as well as using simpler environments with fewer visual elements.

[a]LM: leap motion.

[b]AR: augmented reality.

[c]HMD: head-mounted display.

Conclusions

This study involved 12 participants with dementia who performed 5 different tasks using 5 interactive technologies that were available at the time of this study. As participants used the technologies to perform virtual reality tasks, we identified potential issues, such as assistance provided, comprehension issues, perception issues, interaction issues, and discomfort. We also studied how the patient's profile would affect performance in those different tasks and technologies. Finally, we provided a set of recommendations for the selection, use, and design of virtual tasks for these technologies. Our main findings show significant effects of technology on performance regarding comprehension, interaction, and discomfort.

Overall, the participants were able to complete all tasks using all technologies. However, a clear outcome of the study is that there is no absolute best technology for people with dementia, but this is both task- and patient-profile–dependent. In general, the use of technologies that require direct interaction is advisable, given that cognitive performance gradually declines in people with dementia, as it relies on fewer cognitive resources than indirect interaction devices. We observed that cognitive skills, as assessed by the MMSE test, influenced participants' perception, comprehension, and interaction and required more assistance. In addition, schooling is also a factor to be considered; the lack of experience and exposure to such technologies can lead to confusion and anxiety, which interferes with user experience.

A cost-effectiveness analysis comparing price and issues identified in all technologies suggests that the best tradeoff between performance and cost is achieved with the mouse, the most effective technology is HMD, and the most expensive one is AR. Through these insights, this study provides newer insights for health professionals, informal caregivers, and engineers regarding the use and design of novel technologies for people with dementia to (1) maximize their success in using such technologies to fulfill virtual tasks and (2) safeguard their psychological well-being. The findings of this study and the proposed guidelines are being implemented on a set of SGs for cognitive stimulation, which explores the potential benefits of music and reminiscence-related approaches in people with dementia.

To conclude, the participants in this study were able to handle the technologies to complete virtual tasks. Interestingly, the overall success in using technologies by people with dementia depends on different variables, such as patient profile, type of task, and interaction modality. Our study provides a quantitative analysis that contributes toward a better understanding of the complex relationships among these factors. Finally, by translating our findings into a set of guidelines, we hope to facilitate technological interventions and to enhance the user experience of people with dementia when performing virtual tasks with *out-of-the-shelf* technologies.

Limitations

This study has some limitations. We had a small number of participants, and not all participants interacted with all technologies. Consequently, if we applied Bonferroni correction for multiple comparisons, statistical significances during post hoc analysis do not remain. Hence, a larger sample size would have provided higher statistical power for the analysis. In addition, having a control group of healthy age- and sex-matched participants would have been informative to discriminate age- and dementia-related issues, such as perception problems. Future studies should consider adding a control group to draw additional conclusions regarding the usage of *out-of-the-shelf* technologies to perform virtual reality tasks. Nevertheless, adding a control group presented some challenges.

First, we interacted with a population that cannot adequately express themselves in the same way as healthy elderly people. Therefore, we had to use very time-consuming methodologies, such as independent annotation of hours of video recordings, categorization, and extraction of data, so that a quantitative analysis could be performed. Consequently, we would have to use the same methodology as the control group (that would not require it), making it not feasible for us, given the time needed and available human resources. Second, even if we did so, our experience tells us that the 2 groups would not be directly comparable even if performing the same activities because people with dementia required constant stimulation and assistance by researchers and health professionals to understand and perform the tasks. Third, the level of autonomy of people with dementia in performing the activities is not comparable with that of a healthy old adult.

In addition, when performing the cost-effectiveness analysis, we considered different approaches, such as the normalization of costs and issues. However, we realized that the resulting values obscured the actual relation to either cost or actual issues, making it very difficult to interpret. Moreover, we considered performing an *issues per euro* analysis; however, such an approach was also problematic because the metric favored expensive equipment. That is, the more expensive the equipment, the less the issues and cost ratio. Similarly, very cheap equipment, such as the mouse, always presents a very high (comparatively) issue/cost ratio. Therefore, in the cost-effectiveness analysis, we gave the same weight to issues and cost because (1) it is fairer to compare and (2) easier to interpret.

Moreover, the use of assessment tools should be considered for additional qualitative data analysis, such as the Individually Prioritized Problem Assessment and the Psychosocial Impact of Assistive Devices Scale, to evaluate how technology impacts the daily life of people with dementia. Finally, some video recordings were corrupted, and, although we also used written notes, some level of detail may have been lost.

Acknowledgments

The authors would like to thank the Portuguese Alzheimer Association, health professionals, and participants for their availability and motivation during the authors' visit. In addition, the authors would like to thank José Camacho for allowing them to use the virtual forest scenario during the experimental trial. This project was partially supported by the Portuguese Foundation for Science and Technology under the project NOVA-LINCS-NOVA Laboratory for Computer Science and Informatics-(PEest/UID/CEC/04516/2019), BioVisualSpeech (CMUP-ERI/TIC/0033/2014 and ARDITI-Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação-under the scope of the project M1420-09-5369-FSE-000001 Ph.D. The study was supported as well by the projects LARSyS UID/EEA/50009/2019, by BRaNT-Belief Bevision applied to Neurorehabilitation Therapy (reference number PTDC/CCICOM/30990/2017) financed by FCT- Fundação para a Ciência e a

Tecnologia, and also by MACBIOIDI: Promoting cohesion of Macaronesian regions through a common ICT platform for biomedical R-D-i (INTERREG program MAC/1.1.b/098).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation between participants profile performance and technology.
[DOCX File , 21 KB - games_v8i3e17565_app1.docx]

Multimedia Appendix 2
Scatter plots.
[DOCX File , 101 KB - games_v8i3e17565_app2.docx]

Multimedia Appendix 3
Correlations between patient profile and overall technology.
[DOCX File , 13 KB - games_v8i3e17565_app3.docx]

Multimedia Appendix 4
Correlation between patient profile and task.
[DOCX File , 14 KB - games_v8i3e17565_app4.docx]

Multimedia Appendix 5
Correlation between participants profile and performance with different technologies.
[DOCX File , 14 KB - games_v8i3e17565_app5.docx]

Multimedia Appendix 6
Correlation between participants profile and individual performance domain.
[DOCX File , 14 KB - games_v8i3e17565_app6.docx]

Multimedia Appendix 7
Relationship between performance issues that emerged while performing tasks with different technologies and purchase costs of technology. The price (euros) was that of the equipment when it was purchased.
[PNG File , 41 KB - games_v8i3e17565_app7.png]

References

1. Schell J. The Art of Game Design: A Book of Lenses. New York, USA: CRC Press; 2014.
2. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
3. Ribeiro J, Almeida JE, Rossetti RJ, Coelho A, Coelho AL. Using Serious Games to Train Evacuation Behaviour. In: 7th Iberian Conference on Information Systems and Technologies. 2012 Presented at: CISTI'12; 20-23 June, 2012; Madrid, Spain URL: https://www.researchgate.net/publication/227912263_Using_Serious_Games_to_Train_Evacuation_Behaviour
4. Bul KC, Franken IH, van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, et al. Development and user satisfaction of 'plan-it commander,' a serious game for children with ADHD. Games Health J 2015 Dec;4(6):502-512. [doi: 10.1089/g4h.2015.0021] [Medline: 26325247]
5. Tate R, Haritatos J, Cole S. HopeLab's approach to re-mission. Int J Learn Media 2009;1(1):29-35. [doi: 10.1162/ijlm.2009.0003]
6. Rizzo A, Pair J, Graap K. A virtual reality exposure therapy application for Iraq war military personnel with post traumatic stress disorder: from training to toy to treatment. NATO Security through Science Series E Human and Societal Dynamics 2006;6:235. [doi: 10.1109/vr.2006.23]
7. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22(1):77-89. [doi: 10.3109/14639239709089835] [Medline: 9183781]
8. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertain Comput 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
9. Saposnik G, Teasell R, Mamdani M, Hall J, McIlroy W, Cheung D, Stroke Outcome Research Canada (SORCan) Working Group. Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical

trial and proof of principle. Stroke 2010 Jul;41(7):1477-1484 [FREE Full text] [doi: 10.1161/STROKEAHA.110.584979] [Medline: 20508185]
10. Assad O, Hermann R, Lilla D. Motion-Based Games for Parkinson's Disease Patients. In: International Conference on Entertainment Computing. 2011 Presented at: ICEC'11; October 5-8, 2011; Vancouver, BC, Canada. [doi: 10.1007/978-3-642-24500-8_6]
11. Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L, et al. Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. J Child Psychol Psychiatry 2010 Aug;51(8):944-952. [doi: 10.1111/j.1469-7610.2010.02258.x] [Medline: 20646129]
12. McCallum S, Boletsis C. Dementia Games: A Literature Review of Dementia-Related Serious Games. In: International Conference on Serious Games Development and Applications. 2013 Presented at: SGDA'13; September 25-27, 2013; Trondheim, Norway. [doi: 10.1007/978-3-642-40790-1_2]
13. Sachdev PS, Blacker D, Blazer DG, Ganguli M, Jeste DV, Paulsen JS, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol 2014 Nov;10(11):634-642. [doi: 10.1038/nrneurol.2014.181] [Medline: 25266297]
14. Alain L. Alzheimer's Disease: A Review. New York, USA: Springer; 2018.
15. -. Alzheimer's disease facts and figures. Alzheimer Dementia 2017 Apr;13(4):325-373. [doi: 10.1016/j.jalz.2017.02.001]
16. Calderon J, Perry RJ, Erzinclioglu SW, Berrios GE, Dening TR, Hodges JR. Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. J Neurol Neurosurg Psychiatry 2001 Feb;70(2):157-164 [FREE Full text] [doi: 10.1136/jnnp.70.2.157] [Medline: 11160462]
17. Webster KE, Merory JR, Wittwer JE. Gait variability in community dwelling adults with Alzheimer disease. Alzheimer Dis Assoc Disord 2006;20(1):37-40. [doi: 10.1097/01.wad.0000201849.75578.de] [Medline: 16493234]
18. Lyketsos CG, Lopez O, Jones B, Fitzpatrick AL, Breitner J, DeKosky S. Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. J Am Med Assoc 2002 Sep 25;288(12):1475-1483. [doi: 10.1001/jama.288.12.1475] [Medline: 12243634]
19. Román GC, Erkinjuntti T, Wallin A, Pantoni L, Chui HC. Subcortical ischaemic vascular dementia. Lancet Neurol 2002 Nov;1(7):426-436. [doi: 10.1016/s1474-4422(02)00190-4] [Medline: 12849365]
20. McKeith IG. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. J Alzheimers Dis 2006;9(3 Suppl):417-423. [doi: 10.3233/jad-2006-9s347] [Medline: 16914880]
21. Englund B, Brun A, Gustafson L. Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester groups. J Neurol Neurosurg Psychiatry 1994 Apr;57(4):416-418 [FREE Full text] [doi: 10.1136/jnnp.57.4.416] [Medline: 8163988]
22. Korczyn AD. Mixed dementia--the most common cause of dementia. Ann N Y Acad Sci 2002 Nov;977:129-134. [doi: 10.1111/j.1749-6632.2002.tb04807.x] [Medline: 12480742]
23. Monin JK, Schulz R. Interpersonal effects of suffering in older adult caregiving relationships. Psychol Aging 2009 Sep;24(3):681-695 [FREE Full text] [doi: 10.1037/a0016355] [Medline: 19739924]
24. Wimo A, Winblad B, Jönsson L. The worldwide societal costs of dementia: estimates for 2009. Alzheimers Dement 2010 Mar;6(2):98-103. [doi: 10.1016/j.jalz.2010.01.010] [Medline: 20298969]
25. Wimo A, Guerchet M, Ali G, Wu Y, Prina AM, Winblad B, et al. The worldwide costs of dementia 2015 and comparisons with 2010. Alzheimers Dement 2017 Jan;13(1):1-7 [FREE Full text] [doi: 10.1016/j.jalz.2016.07.150] [Medline: 27583652]
26. Wollen KA. Alzheimer's disease: the pros and cons of pharmaceutical, nutritional, botanical, and stimulatory therapies, with a discussion of treatment strategies from the perspective of patients and practitioners. Altern Med Rev 2010 Sep;15(3):223-244 [FREE Full text] [Medline: 21155625]
27. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004 Aug;3(8):711-715. [doi: 10.1038/nrd1470] [Medline: 15286737]
28. Robert PH, König A, Amieva H, Andrieu S, Bremond F, Bullock R, et al. Recommendations for the use of serious games in people with Alzheimer's disease, related disorders and frailty. Front Aging Neurosci 2014;6:54 [FREE Full text] [doi: 10.3389/fnagi.2014.00054] [Medline: 24715864]
29. Zheng J, Chen X, Yu P. Game-based interventions and their impact on dementia: a narrative review. Australas Psychiatry 2017 Dec;25(6):562-565. [doi: 10.1177/1039856217726686] [Medline: 28856903]
30. Tarnanas I, Tsolaki M, Nef T, Müri RM, Mosimann UP. Can a novel computerized cognitive screening test provide additional information for early detection of Alzheimer's disease? Alzheimers Dement 2014 Nov;10(6):790-798. [doi: 10.1016/j.jalz.2014.01.002] [Medline: 24656838]
31. Tong T, Chan J, Chignell M. Serious Games for Dementia. In: Companion: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia. [doi: 10.1145/3041021.3054930]
32. Lancaster C, Koychev I, Blane J, Chinner A, Wolters L, Hinds C. Evaluating the feasibility of frequent cognitive assessment using the Mezurio smartphone app: observational and interview study in adults with elevated dementia risk. JMIR Mhealth Uhealth 2020 Apr 2;8(4):e16142 [FREE Full text] [doi: 10.2196/16142] [Medline: 32238339]

33. D'Onofrio G, Sancarlo D, Ricciardi F, Panza F, Seripa D, Cavallo F, et al. Information and communication technologies for the activities of daily living in older patients with dementia: a systematic review. J Alzheimers Dis 2017;57(3):927-935. [doi: 10.3233/JAD-161145] [Medline: 28304297]
34. Foloppe DA, Richard P, Yamaguchi T, Etcharry-Bouyx F, Allain P. The potential of virtual reality-based training to enhance the functional autonomy of Alzheimer's disease patients in cooking activities: a single case study. Neuropsychol Rehabil 2018 Jul;28(5):709-733. [doi: 10.1080/09602011.2015.1094394] [Medline: 26480838]
35. Benveniste S, Jouvelot P, Pin B, Péquignot R. The MINWii project: renarcissization of patients suffering from Alzheimer's disease through video game-based music therapy. Entertain Comput 2012 Dec;3(4):111-120. [doi: 10.1016/j.entcom.2011.12.004]
36. Morel A, Bormans K, Rombouts K. Memory Palaces to Improve Quality of Life in Dementia. In: Conference on Raising Awareness for the Societal and Environmental Role of Engineering and (Re)Training Engineers for Participatory Design. 2015 Presented at: Engineering4Society'15; June 18-19, 2015; Leuven, Belgium. [doi: 10.1109/engineering4society.2015.7177904]
37. Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, et al. Categorizing health outcomes and efficacy of mhealth apps for persons with cognitive impairment: a systematic review. J Med Internet Res 2017 Aug 30;19(8):e301 [FREE Full text] [doi: 10.2196/jmir.7814] [Medline: 28855146]
38. Laird EA, Ryan A, McCauley C, Bond RB, Mulvenna MD, Curran KJ, et al. Using mobile technology to provide personalized reminiscence for people living with dementia and their carers: appraisal of outcomes from a quasi-experimental study. JMIR Ment Health 2018 Sep 11;5(3):e57 [FREE Full text] [doi: 10.2196/mental.9684] [Medline: 30206053]
39. Joddrell P, Astell AJ. Studies involving people with dementia and touchscreen technology: a literature review. JMIR Rehabil Assist Technol 2016 Nov 4;3(2):e10 [FREE Full text] [doi: 10.2196/rehab.5788] [Medline: 28582254]
40. Bejan A, Wieland M, Murko P, Kunze C. A Virtual Environment Gesture Interaction System for People with Dementia. In: Companion: Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems. 2018 Presented at: DIS'18; June 9-13, 2018; Hong Kong, China. [doi: 10.1145/3197391.3205440]
41. McLaughlin AC, Rogers WA, Fisk AD. Using direct and indirect input devices: attention demands and age-related differences. ACM Trans Comput Hum Interact 2009 Apr;16(1):1-15 [FREE Full text] [doi: 10.1145/1502800.1502802] [Medline: 22563232]
42. Winn B. The design, play, and experience framework. In: Handbook of Research on Effective Electronic Gaming in Education. Hershey, Pennsylvania: IGI Global; 2009.
43. Vallejo V, Tarnanas I, Yamaguchi T, Tsukagoshi T, Yasuda R, Müri1 R, et al. Usability assessment of natural user interfaces during serious games: adjustments for dementia intervention. J Pain Manag 2014;9(3):333-339 [FREE Full text]
44. Alvseike H, Brønnick K. Feasibility of the iPad as a hub for smart house technology in the elderly; effects of cognition, self-efficacy, and technology experience. J Multidiscip Healthc 2012;5:299-306 [FREE Full text] [doi: 10.2147/JMDH.S35344] [Medline: 23226024]
45. Astell A. REAFF - A Framework for Developing Technology to Address the Needs of People With Dementia. In: Proceedings of First International Workshop. 2009 Presented at: RSW'09; May 11, 2009; Aachen, Germany URL: https://www.researchgate.net/publication/289804258_REAFF_-_A_framework_for_developing_technology_to_address_the_needs_of_people_with_dementia
46. Astell A. Creating technologies with people who have dementia. In: Perspectives on Human-Computer Interaction Research with Older People. New York, USA: Springer; 2019.
47. Krohn S, Tromp J, Quinque EM, Belger J, Klotzsche F, Rekers S, et al. Multidimensional evaluation of virtual reality paradigms in clinical neuropsychology: application of the VR-check framework. J Med Internet Res 2020 Apr 27;22(4):e16724 [FREE Full text] [doi: 10.2196/16724] [Medline: 32338614]
48. Tomé R, Pereira J, Oliveira M. Using Serious Games for Cognitive Disabilities. In: International Conference on Serious Games Development and Applications. 2014 Presented at: SGDA'14; October 9-10, 2014; Berlin, Germany. [doi: 10.1007/978-3-319-11623-5_4]
49. Bouchard B, Imbeault F, Bouzouane A. Developing Serious Games Specifically Adapted to People Suffering from Alzheimer. In: International Conference on Serious Games Development and Applications. 2012 Presented at: SGDA'12; September 26-29, 2012; Bremen, Germany. [doi: 10.1007/978-3-642-33687-4_21]
50. Tziraki C, Berenbaum R, Gross D, Abikhzer J, Ben-David BM. Designing serious computer games for people with moderate and advanced dementia: interdisciplinary theory-driven pilot study. JMIR Serious Games 2017 Jul 31;5(3):e16 [FREE Full text] [doi: 10.2196/games.6514] [Medline: 28760730]
51. Meiland F, Innes A, Mountain G, Robinson L, van der Roest H, García-Casal JA, et al. Technologies to support community-dwelling persons with dementia: a position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1 [FREE Full text] [doi: 10.2196/rehab.6376] [Medline: 28582262]
52. Thordardottir B, Fänge AM, Lethin C, Gatta DR, Chiatti C. Acceptance and use of innovative assistive technologies among people with cognitive impairment and their caregivers: a systematic review. Biomed Res Int 2019;2019:9196729 [FREE Full text] [doi: 10.1155/2019/9196729] [Medline: 30956989]

53. Broady T, Chan A, Caputi P. Comparison of older and younger adults' attitudes towards and abilities with computers: implications for training and learning. Br J Educ Technol 2010;41(3):485. [doi: 10.1111/j.1467-8535.2008.00914.x]
54. Laguna K, Babcock RL. Computer anxiety in young and older adults: implications for human-computer interactions in older populations. Comput Hum Behav 1997 Aug;13(3):317-326. [doi: 10.1016/s0747-5632(97)00012-5]
55. Mondellini M, Arlati S, Greci L, Ferrigno G, Sacco M. Sense of Presence and Cybersickness While Cycling in Virtual Environments: Their Contribution to Subjective Experience. In: International Conference on Augmented Reality, Virtual Reality and Computer Graphics. 2018 Presented at: AVR'18; June 24-27, 2018; Otranto, Italy. [doi: 10.1007/978-3-319-95270-3_1]
56. Hackner E, Lankes M. Mindtraining: Playful Interaction Techniques for People with Dementia. In: International Conference on Entertainment Computing. 2016 Presented at: ICEC'16; September 28-30, 2016; Wien, Austria. [doi: 10.1007/978-3-319-46100-7_21]
57. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc 1992 Sep;40(9):922-935. [doi: 10.1111/j.1532-5415.1992.tb01992.x] [Medline: 1512391]
58. Guna J, Jakus G, Pogačnik M, Tomažič S, Sodnik J. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors (Basel) 2014 Feb 21;14(2):3702-3720 [FREE Full text] [doi: 10.3390/s140203702] [Medline: 24566635]
59. Tools. The NeuroRehabLab. URL: https://neurorehabilitation.m-iti.org/tools/ [accessed 2020-04-05]
60. Jam Session. Steam Community. URL: https://steamcommunity.com/sharedfiles/filedetails/?id=928174467 [accessed 2020-04-05]
61. Blocks. Leap Motion Gallery. URL: https://gallery.leapmotion.com/blocks/ [accessed 2020-03-22]
62. Virtual Piano For Beginners. Leap Motion Gallery. URL: https://gallery.leapmotion.com/virtual-piano-for-beginners/ [accessed 2020-03-22]
63. Ghost Ship. Welcome to Steam. URL: https://store.steampowered.com/app/668180/Ghost_Ship/ [accessed 2020-03-22]

Abbreviations

AR: augmented reality
HMD: head-mounted display
LM: leap motion
MMSE: Mini-Mental State Examination
REAFF: responding, enabling, augmenting, failure free
RQs: research questions
SG: serious game
VR: virtual reality

Edited by T Rashid Soron; submitted 20.12.19; peer-reviewed by S Krohn, C Leorin, E Siegmund-Schultze; comments to author 30.01.20; revised version received 15.04.20; accepted 09.06.20; published 05.08.20.

Please cite as:
Andrade Ferreira LD, Ferreira H, Cavaco S, Cameirão M, i Badia SB
User Experience of Interactive Technologies for People With Dementia: Comparative Observational Study
JMIR Serious Games 2020;8(3):e17565
URL: http://games.jmir.org/2020/3/e17565/
doi:10.2196/17565
PMID:32755894

Original Paper

Exploring User Needs in the Development of a Virtual Reality–Based Advanced Life Support Training Platform: Exploratory Usability Study

Nathan Moore[1], MEd, BSN; Soojeong Yoo[2], PhD; Philip Poronnik[3], PhD; Martin Brown[4], MHum; Naseem Ahmadpour[2], PhD

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

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

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

[4]Innovative Technologies, Office of the Vice-Chancellor and Principal Westmead Operations, The University of Sydney, Sydney, Australia

Corresponding Author:
Nathan Moore, MEd, BSN
Research and Education Network
Western Sydney Local Health District
Post Graduate Medical Education Centre Lvl 3 Education Block, Westmead Hospital
Cnr Hawkesbury Rd and Darcy Rd
Westmead, 2145
Australia
Phone: 61 427850889
Email: nathan.moore@health.nsw.gov.au

Abstract

Background: Traditional methods of delivering Advanced Life Support (ALS) training and reaccreditation are resource-intensive and costly. Interactive simulations and gameplay using virtual reality (VR) technology can complement traditional training processes as a cost-effective, engaging, and flexible training tool.

Objective: This exploratory study aimed to determine the specific user needs of clinicians engaging with a new interactive VR ALS simulation (ALS-SimVR) application to inform the ongoing development of such training platforms.

Methods: Semistructured interviews were conducted with experienced clinicians (n=10, median age=40.9 years) following a single playthrough of the application. All clinicians have been directly involved in the delivery of ALS training in both clinical and educational settings (median years of ALS experience=12.4; all had minimal or no VR experience). Interviews were supplemented with an assessment of usability (using heuristic evaluation) and presence.

Results: The ALS-SimVR training app was well received. Thematic analysis of the interviews revealed five main areas of user needs that can inform future design efforts for creating engaging VR training apps: affordances, agency, diverse input modalities, mental models, and advanced roles.

Conclusions: This study was conducted to identify the needs of clinicians engaging with ALS-SimVR. However, our findings revealed broader design considerations that will be crucial in guiding future work in this area. Although aligning the training scenarios with accepted teaching algorithms is important, our findings reveal that improving user experience and engagement requires careful attention to technology-specific issues such as input modalities.

(JMIR Serious Games 2020;8(3):e20797) doi:10.2196/20797

KEYWORDS

virtual reality; Advanced Life Support; Advanced Cardiac Life Support; clinical training; education; serious games

Introduction

Background

Advanced Life Support (ALS) training is a mandatory requirement for all medical staff with critical care exposure and all nursing staff are required to attend as part of the ALS team. The term Advanced Cardiac Life Support (ACLS) is synonymous with ALS, with the latter being the generally accepted vernacular in Australia. An ALS team comprises a number of clinicians with a variety of skills and experiences. For such a team to function at the highest level, it is essential that a skilled clinician is identified to take the role of team leader [1]. The role of ALS team leader requires the clinician to manage a patient who is experiencing an acute deterioration or already in cardiac arrest. This management is largely built upon endorsed algorithms such as the Australian Resuscitation Guidelines [2]. These guidelines are typically applied in a highly stressful and chaotic environment where 6 to 10 clinicians must be directed in the tasks required to manage the patient. Unfortunately, many clinicians may be inadequately prepared to do these tasks [3]. The stress and chaos resulting in deviation from the guidelines may lead to catastrophic outcomes for the patient [4].

Currently, ALS training in many hospitals consists of prereading from a manual outlining the procedure algorithm; quizzes with multiple-choice questions; and face-to-face training days including skills stations, didactic sessions, simulations, and a face-to-face assessment. This assessment is then repeated as per institutional requirements, often on an annual basis.

Initial face-to-face training, refresher training, and reaccreditation are resource-intensive processes requiring high instructor-to-participant ratios and costly simulation equipment [5]. Refresher training can pose challenges for educators and clinicians alike [6]. There are logistical challenges common to face-to-face training and reaccreditation, such as getting clinicians with high clinical demands off the hospital floor as well as coordinating recourses and educator time [7].

During face-to-face training, a significant amount of time is spent exploring the technical elements of the ALS algorithm and management, such as drug dosages and timings, rhythm interpretation, defibrillation timings, differential diagnoses, and more. Although these technical elements are important, the exploration and training of them detracts from time that could be spent usefully on the equally important nontechnical elements of ALS management such as communication, teamwork, and situational awareness [8]. These resource challenges have driven a need moving forward to focus on efficient and scalable high-quality health education [9,10].

Objectives

There is a known and well-represented correlation between reducing the time taken to identify patient deterioration and a marked improvement in patient outcomes in the inpatient setting [11]. As a result, multitiered rapid response systems and mandatory standardized triggers for escalation in patient management have been implemented, leading to earlier identification of patient deterioration [12]. A recent systematic review showed that this early identification has resulted in a pronounced decrease in the number of cardiac arrest calls occurring in inpatient populations [13]. The challenge this presents for clinicians is that cardiac arrest management has now become a low-frequency yet high-stakes event. This in turn means that clinicians have reduced opportunities to perform ALS and consolidate skills. However, when required, they are expected to execute these skills at the highest standard. Even with exposure to these skills in a clinical setting, clinicians require more than practice alone, as further feedback and education are essential [10].

Clinicians and educators must, therefore, explore modalities and technologies that can overcome the barriers of cost, access, and frequency of exposure [14]. Some of the ways of addressing these challenges include simulation-based learning using the latest technologies (eg, virtual reality [VR], which has generated a growing interest in recent years [15]). Studies have demonstrated improved performance when using VR for basic life support training [16], but there is limited information about what users of ALS systems need, specific design considerations, and ways we can improve user experience and engagement with the training [17].

The main objective of this study is to identify key user needs and guidance for designing VR-based educational resources for ALS providers. We achieve this through an exploratory study, using a virtual ALS training module we have created.

VR Technology for ALS Training

VR offers a number of novel capabilities that have promising potential in providing new kinds of support for educators [18,19]. The ability to visualize and interact with 3D representations in real time and visualize the dynamic relationships between variables in a complex environment [20] makes for a compelling use case in the context of ALS training. In addition, VR technology provides a simulated environment that can help users recreate an experience which would otherwise be difficult to have [21]. Finally, VR provides a unique immersive experience that blocks out the physical world, resulting in the user believing or feeling they are present in the VR environment [22].

These affordances of VR make it an intriguing platform to explore for ALS training. There have been a few attempts at developing similar VR-based training such as Health Scholars ACLS Virtual Reality training [23] or dualgood ACLS and ALS training [24]. However, there are a number of limitations associated with using these platforms in our context, including potential costs as a commercial product and a lack of customization of the training modules, which are based on the American Heart Association guidelines [25] and differ from those endorsed in other countries, such as the Australian Resuscitation Guidelines [26].

In addition to system limitations that prevent the scalability of some current ALS trainings in VR, there is currently a knowledge gap about the perspectives and needs of the end users, who should be included in the development of future applications in this field [27]. A user-centered approach would involve a careful and iterative process that takes into account the training requirements, determinants of positive user experience among trainees, and usability of the system to provide a smooth and engaging interaction in the virtual environment. In this study, we describe and explore this approach in the context of ALS training, and specifically in relation to the most challenging role, that of the ALS team leader.

We identified a few challenges for designing VR to supplement training for ALS team leaders, based on our own experiences of delivering this type of training. One challenge is that participants are often unable to recall information available within the prescribed prereading on the face-to-face training days. The concept of learner fatigue and disengagement when faced with a large amount of prereading is well established [28]. The prereading for this program is a 188-page manual intended as an ongoing resource; however, participants are expected to have a basic understanding of the majority of the concepts contained within prior to the program. The level of immersion in VR has been shown to further facilitate information recall presented in the virtual environment compared to a flat screen [29], upon which the training manual could be viewed. This effect can be amplified through creating a realistic setting in VR to enhance information recall [30]. Furthermore, VR enables the user to explore alternate viewpoints in a given scenario [20]. This could be used to enable the user to see the environment from the viewpoint of other team members as well as to review their own performance from alternate viewpoints.

Interactive VR Simulations in Medical Education

Interactive virtual simulations have been utilized in medical education to provide students and clinicians engaging ways to learn skills, abilities, and critical knowledge for real-world applications [31,32]. Immersive virtual environments (VE) have been used in the context of health care education to offer development of technical skills in the areas of surgical training [33] and clinical procedures [34], as well as replicating complex team-based systems such as ALS [31]. Until recently, one of the limitations of these VEs has been the need to have high-powered computers, networks, and tethered head-mounted devices (HMD) to provide high-quality VR. The release of the Oculus Quest (Oculus VR), which has 6 degrees of movement and a relatively low unit cost, has allowed for a standalone portable, powerful, and immersive experience.

A notable example of VR-based ALS training simulation is the app developed by Arizona State University [18], which provides a multiplayer collaborative gaming VE. Although there are clear benefits for practicing as a team in a dynamic ALS setting [35], this requires multiple users to be coordinated to practice at the same time in the same VE. This poses a significant logistic and time challenge and does not allow individuals to rehearse on their own time.

In this study, we describe a single-player interactive simulation that overcomes the limitations mentioned above. We developed a prototype VR app that contains elements of interactive simulation, which we then used to study the factors relevant to users' needs to deliver a positive training experience that is also easy to use. Elements that were of particular importance to the interactivity of the experience include gameplay, a clear purpose, and realistic representation of current environments [36].

Design

In our experience developing several simulation and flipped-classroom supplemental training assets, any solution to supplement ALS training needs to be flexible and meet the requirements of both the educators and clinicians. We identified two primary usage opportunities that could be of value to the identified audience of an interactive VR-based app and used those to create a frame of reference for prototyping our app, named ALS-SimVR.

The first is the opportunity to use the VR app intermittently in a clinical setting. There is often an overlap period of 30 to 45 minutes during handovers that is traditionally allocated to in-service time. Designing for this opportunity means ensuring that the VR app allows for a complete experience for the user within these time frames. This also implies that the ALS-SimVR app should be deployed on a portable device and require little to no setup time.

The second opportunity is the use of the app for preparation or refreshing knowledge during the user's private time (eg, at home or in their office). Similarly, this requires portability and an easy setup process.

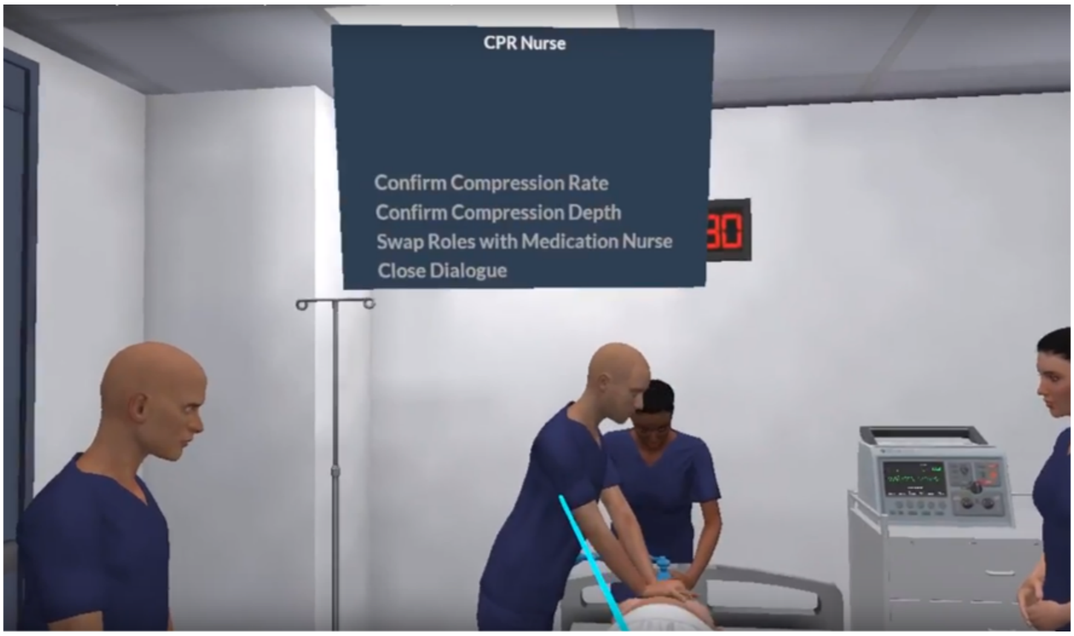
ALS-SimVR Application Walkthrough

The interactive simulation places the user in the position of ALS team leader and allows them to direct the team in the management of a patient in cardiac arrest [37]. The virtual space as seen from the team leader's point of view is shown in Figure 1.

Figure 1. In-game user view displaying the team leader's perspective.

Upon commencement of the scenario, the team leader is presented with a handover from one of the virtual team members, outlining the current clinical situation. The team leader is then free to manage the patient as per their clinical judgement. All management decisions are made by clicking on a virtual team member using the Oculus Go controller via dropdown menus that appear above the virtual team member's head. These decisions are all recorded and represented on the scribe notes within the application, which are available as a reference for the user. Throughout the scenario, realistic prompts are provided by the virtual team members, such as "difficulty ventilating the patient" or "nearing 2 minutes of CPR," all of which are generalizable clinical cues which replicate real-world team function (Figure 2).

Figure 2. Interactions with virtual team members.



Vital tasks such as defibrillation, drug and fluid administration, compression/ventilation ratio and rate, role swapping, intravenous access, and airway management are all available as decisions and are replicated realistically in the scenario. The user is free to make decisions at any time they choose, whether the decision is correct or incorrect. Those choices and timings form the assessment provided at the culmination of the scenario.

The scenario within ALS-SimVR mirrors the face-to-face ALS training at Westmead Hospital in New South Wales, Australia. The base scenario was designed in consultation with senior clinicians directly involved in cardiac arrest management and reviewed to realistically replicate a common presentation of cardiac arrest. The scenario mimics pathology and decisions that are generalizable across Australia.

Overview of the ALS-SimVR App Architecture

The application was created initially for deployment on the Oculus GO. The portable nature of these devices means that clinicians can practice at their own pace prior to attending an ALS program session or annual reaccreditation.

ALS-SimVR was built using the Unreal game engine [38]. This choice was made as it is considered to have textures and editing capacity and provides a finished product of a high visual standard. At the heart of the ALS-SimVR app is the replication of the complex decision-making processes. The actions are broken down into small segments of code that populate decision trees. Using Unreal visual scripting, these actions and descriptions were formed into nodes. These nodes were then compiled together to make blueprints that convert user decisions into game-based actions. The use of blueprints allows for actions containing code, audio, animation, and facial motion capture to easily be used in other sequences and scenarios. Running concurrently with the application were analytics that recorded all decisions and interactions. These analytics allow for the generation of the user feedback document provided upon the completion of the scenario. The analytics also populated the scribe notes, which were available for reference upon user request during the scenario. This is a vital resource utilized by clinicians during ALS events.

Methods

In total, 10 clinicians with ALS experience in both clinical and educational settings were recruited to provide expert input. The research was approved by the University of Sydney Human Ethics Committee (2016/089). All participants completed written informed consent forms.

At the start of the experiment, participants were given a brief overview of the study and instructed on how to use the Oculus GO headset. Participants then performed a playthrough of the cardiac arrest scenario and managed the patient as they would in a real clinical setting. Their sessions were recorded using the Oculus livestream feature and ran for approximately 10 minutes.

Upon completion of the scenario, a semistructured interview was conducted, which was audio-recorded with the participant's permission. Each interview lasted approximately 20 minutes. Participants were first asked about their immediate overall impression and experience. Participants were then asked about the usability of the application and their own performance. These usability questions asked the participant for their thoughts about the 12 heuristics items in Sutcliffe and Gault's heuristic evaluation of VR apps [39], which were read out to the participant to ensure a consistent but broad coverage of specific performance aspects. If the participant's response was vague, the interviewer asked for further clarity. The heuristic items are the following: natural engagement, compatibility with the use's tasks and domain, natural expression of action, close coordination of actions and representation, realistic feedback, faithful viewpoints, navigation and orientation support, clear entry and exit points, consistent departures, support for learning, clear turn-taking, and sense of presence. Next, the participants rated the severity of performance issues relevant to each item on a 4-point scale (Severe, Annoying, Distracting, or Inconvenient) as per the standard questionnaire [39].

Finally, participants provided any further comments before completing the Université du Québec en Outaouais Revised Presence Questionnaire (PQ) with 24 items, each rated on a 7-point scale [40]. In the context of this study, we excluded items 23 and 24 as they were concerned with haptic experiences that did not apply to ALS-SimVR. The 22 remaining items are classified into 7 presence factors in the PQ: realism, possibility to act, quality of interface, possibility to examine, self-evaluation of performance, sounds, and haptics. Each item is formulated as a question (eg, How much were you able to control events?) with the 7-point scale ranging from the lowest score (1=Not at all) to the highest (7=Completely).

Results

Overview

Participants (n=6 female, n=4 male) had an average age of 40.9 (SD 9.48) years, and an average of 12.4 (SD 6.93) years of experience providing or teaching ALS. All participants had minimal to no previous experience in using VR in either a personal or education capacity. All audio interviews were transcribed by a blinded researcher. Two researchers (one expert in teaching ALS in the clinical setting and another a human-computer interaction researcher) worked together and followed an inductive (or bottom-up) approach to the analysis of qualitative feedback received from participants [41]. No assumptions were made about a coding system and instead participant statements were clustered based on similarity in content, after which higher level themes were identified and labelled. The analysis aimed to identify the needs and expectations of the participants relative to their experience with the ALS-SimVR. The average ratings for usability and presence were then calculated to reinforce the statements provided in the interviews.

Interviews

A number of insights were derived from the participant statements, which were classified into five themes summarized in Table 1, supplemented with brief descriptions and subthemes when relevant. Each of the five themes are discussed next.

Table 1. User needs groupings and exemplar statements.

User needs theme	Exemplar statements
Affordances	
Realistic tasks	Common clinical tasks should be available for completion in a realistic manner
Visibility	Clear visible assets aligned with environmental orientation
Completion	Clear commencement and completion prompt to task
Accessibility	Clarity as to how commands are given and accessed
Agency	The environment providing opportunity to control workflows autonomously and make choices that align with prior experiences, such as multitasking
Diverse input modalities	The environment replicates natural input modalities such as issuing commands verbally
Mental models	The environment design and prompts align with how the clinical environment operates (eg, 2 minutes of cardiopulmonary resuscitation completion)
Advanced roles	The ability to manage tasks at an acceptable clinical standard

Affordances

Overview

Affordances often represent relational approaches to the way people interact with technologies [42]. In relation to ALS-SimVR, we specify affordances as users' concern for the availability of action within the virtual environment. The participants in our study noted four specific affordances that needed to be addressed through design. These are detailed below.

Realistic Tasks

Participants identified a need for the ability to do the tasks in a virtual environment in a realistic manner, replicating the clinical environment. Although the app was designed to recreate the majority of the initial requirements for ALS management, some participants (in particular, more senior clinicians) were not able to complete the tasks in a manner that mirrored real clinical management. For example, the programming constraint in the ALS-SimVR did not allow the participants to return to their previous decisions and change what they had previously done, which is typical in patient management.

Visibility

Participants highlighted the need for visibility of actors and objects they needed to interact with in the virtual environment. For instance, the positioning of the nurse scribe avatar within the virtual environment was visually inaccessible to the participants (the avatar was slightly behind their right shoulder) and it took a while for participants to notice the scribe was there. This resulted in some participants missing out on some of the functionality of the app. However, once the participants had oriented themselves, they reported the positioning of the team members accessible and "good." A design recommendation emerging from this need is to allow more time within future solutions of this type to allow users to explore their surroundings prior to commencing a scenario. Another example of visibility issues that emerged in the study was that the three degrees of freedom and resolution limits of the VR headset resulted in some participants being unable to read or clearly visualize some assets within the environment.

Completion

Knowing when all tasks were completed and the scenario had actually ended was an affordance issue identified by some participants. Upon completion of the scenario, the participant was presented with a large report document within the app titled "Results." This provided them with feedback on their performance, including compliance with the ALS algorithm, defibrillation, and intervention accuracy. The report appeared in the center of their vision, between the patient and the participant. Some of the participants chose to immediately close the document without reading it and continued with the scenario. Due to the design of the app, they were not able to find and review that document at a later time, which seemed to negatively impact their experience.

Accessibility

Participants highlighted the need for greater orientation in how to start "playing the game." One participant commented "I needed guidance to know how to play the game." Additionally, participants wanted clarity around some design elements that they could not identify affordances for. For instance, participants did not realize how an interactive text box provided the available interactions or tasks and how they could reopen them. Participants expressed their preferences for interface accessibility including readability of design elements (eg, larger text size and clickable elements), visibility of design elements (eg, a more comfortable distance to a pathology report that was positioned too close), or flexibility of actions (eg, flipping through a chart). These were departures from the realities of managing patients in the clinical setting and were consequently deemed "annoying" by participants.

Agency

Participants wanted the opportunity and control to do things autonomously, based on their personal experiences. In the context of health care, this poses challenges as the complex nature of managing unwell patients means clinicians make different choices depending on their seniority, experience, and education. This requires a level of agency that was not supported in ALS-SimVR. Additionally, having the agency to make such choices may require multitasking, which the app did not enable.

Participants wanted the capacity to plan ahead and even allocate tasks to different team members.

Diverse Input Modalities

Participants discussed their desire for a diverse set of inputs to the device, similar to real life, for instance by using voice commands in addition to using the hand controller. One participant stated, “I just didn’t like not being able to talk” and “I wanted to talk to the team to ask for updates.” This particular activity seemed divisive among participants with some expressing an absolute need for it to reflect the real nature of work and others stating this was not a problem for them.

Mental Models

Overall, 3 participants identified instances where the actions they took within the virtual environment did not align with the approach they would take in the clinical environment (ie, their preconceived mental models). Clinical decisions made within the ALS-SimVR are recorded and analyzed to form an assessment of user performance at the completion of the scenario. The way this is programmed within the app meant that once those decisions are made, they cannot be undone or revisited. This created a disconnect for a number of participants whose mental models or real-world practice would entail frequent reassessment of their patient and decisions to escalate the situation accordingly.

Advanced Roles

A number of additional roles in the ALS-SimVR scenario were suggested by the participants to create new interpersonal interactions within the app, for example to access the patient assessment report by asking someone on the team instead of clicking on the patient. One participant noted that the scenario “didn’t have extra people.” Another reported the “need to practice with team members not knowing their roles perfectly” is a constant challenge for real-life implementation of this training. Other suggestions included advanced roles (through avatars) for cardiopulmonary resuscitation and to demonstrate listening to the chest.

Usability Heuristics

The participants rated the severity of each heuristic item [39]. Two items (Item 7 – Navigation and orientation support, and Item 11 – Clear turn taking) were eliminated from the analysis as they were not relevant to this application. The ALS-SimVR is a single player experience that did not allow the user to move in their environment and did not involve turn taking. The results of these ratings are summarized in Table 2. The number of issues reported for each item and their corresponding severity was calculated. For instance, for the first item (natural engagement), we received 2 statements implying a “severe” issue, 5 statements implying an “annoying” issue, 8 statements implying “distracting” issues, and 11 statements implying “inconvenient” issues. A total score was then calculated for each item by adding the total number of reported issues.

Overall, 3 items clearly had the highest usability issues: Item 1 – Natural engagement (with a total score of 26), Item 3 – Natural expression of action (with a total score of 31), and Item 12 – Sense of presence (with a total score of 15). This is an indication of areas of priority for improving the usability of ALS-SimVR.

Table 2. Results of heuristic responses, noting the number of statements relevant to each item and the corresponding severity ratings.

Heuristic item	Severity rating				Total/item
	Severe	Annoying	Distracting	Inconvenient	
Natural engagement	2	5	8	11	26
Compatibility with users’ task and domain	1	—[a]	3	5	9
Natural expression of action	2	3	15	11	31
Close coordination of action and representation	—	—	2	1	3
Realistic feedback	—	—	—	2	2
Faithful viewpoints	1	1	—	2	4
Navigation and orientation support	—	—	—	—	N/A[b]
Clear entry and exit points	—	3	1	3	7
Consistent departures	—	-	1	1	2
Support for learning	1	2	3	5	11
Clear turn taking	—	—	—	—	N/A
Sense of presence	1	—	4	10	15

[a]Not available.

[b]N/A: not applicable.

Presence

The participant responses to the relevant 22 items on the PQ were compiled. The standard scoring system classifies the responses into 7 themes: realism, possibility to act, quality of interface, possibility to examine, self-evaluation of performance, sounds, and haptic (not present in this study). The rating scores received for all items relevant to each theme were averaged and then summed. The average rating scores and standard deviations of the 7 themes are presented in Table 3.

The findings suggest realism received the highest score (average score 35, SD 10.1), indicating the success of ALS-SimVR in replicating a realistic experience and potentially resulting in an enhanced sense of presence in the virtual environment. Conversely, self-evaluation of performance received the lowest score (average score 10.6, SD 3.3), indicating a poor experience of self-evaluation, which diminishes the participant's sense of presence in the virtual environment.

Table 3. The average rating scores on the 7 Presence Questionnaire items indicate the participants' experience of their presence in ALS-SimVR, a virtual reality Advanced Life Support training application.

Presence Questionnaire item	Average (SD)
Realism	35.0 (10.1)
Possibility to act	21.9 (5.6)
Quality of interface	13.1 (5.0)
Possibility to examine	14.0 (3.9)
Self-evaluation of performance	10.6 (3.3)
Sounds	16.5 (5.4)
Haptics	N/A[a]

[a]N/A: not applicable.

Discussion

Principal Findings

Our findings provide insights as to how VR applications could be improved to better address the needs of clinicians in the context of ALS training. There is evidence that the use of virtual patients can be beneficial to increasing the engagement of clinicians [43]; however, the evidence on how that can be achieved is limited. Our study explored specific user needs to improve their engagement with the interactive simulation training app. We conducted usability testing and interviews with 10 expert participants with experience in ALS. Their feedback revealed important consideration for the ALS-SimVR app in five categories that stipulate how VR training should be situated to support the learning experience: affordances, agency, diverse input modalities, mental models, and advanced roles. These provide detailed design directions for improving ALS-SimVR and developing any application of this type in the future.

The usability testing revealed higher scores for natural expression of action, natural engagement, and sense of presence, indicating areas of improvement. Lower scores for usability items such as consistent departures, realistic feedback, as well as close coordination of action, representation, and faithful viewpoints, indicate areas in which ALS-SimVR performed well. The findings from the PQ complemented, gave scope and, in some instances, reinforced some of the findings of the heuristic-based usability testing and uncovered additional user perspectives. For instance, realism and possibility of actions were scored high in the PQ, which can be attributed to the natural engagement and expression of actions mentioned in the usability tests. These are discussed further in this section. The interview findings revealed important avenues for further investigation into the design of future applications in this field, especially in relation to what affordances should be prioritized in design and potential challenges to user engagement.

Clinical Training Through VR: Situating Affordances in the Learning Experience

It is clear from our findings that any VR application for health care training must capture accurate and realistic affordances. Previous evidence linked realism to successful VR training, for instance in surgical training [44]. Our assessment of the ALS-SimVR yielded mixed results.

On the one hand, realism was scored high in the PQ but on the other hand, a relatively high number of presence issues were identified through the heuristic usability testing, with most issues linked to natural expression of actions. Clear opportunities for future development, based on collected feedback, pointed to the need for clarity and accessibility in problem-based learning [45], which underlined the training experience in ALS-SimVR.

The clinicians who participated in our study desired an educational package that allows them to function as they would in the clinical environment in a high-fidelity, realistic, and engaging way. This remains an ongoing challenge in the design of VR-based health care training simulations [19] due to the complex management choices and decision making required to care for deteriorating patients. Our findings suggest clear directions for addressing some of these challenges.

We propose that designing for realistic affordances should focus on "realistic actions" afforded to trainees in the virtual space. For instance, the interactive VR simulation should allow users to perform tasks in a realistic manner, to have good visibility, a clear beginning and end to tasks, and the ability to give commands. The latter, in our experience, is more challenging to achieve. This is also clear from the notable number of issues relevant to natural engagement (see Table 2), as our participants demanded the ability to walk around the virtual environment. This mirrors real-world ALS training where, although the team leader stands at the foot of the patient bed, they often move around to get a better view. This was not afforded by the ALS-SimVR hardware, which at the time only offered three degrees of movement.

Challenges and Opportunities in Interactive Medical VR Simulations

The thematic analysis of the interview results highlighted interaction challenges for users. Our findings suggest that diversifying the input modalities can be a remarkable opportunity to improve the user experience in interactive VR simulations. This is evident from the low score given to the sound item on the PQ (Table 3). A contentious issue among participants was the lack of voice control. A number of participants stated that voice control would not be necessary because the visual communication through the text boxes within the app was clear and worked for them, and they felt they could easily "fill in the gaps." Other participants, however, desired the opportunity to use voice as an input modality.

While this was an important issue, the literature provides limited discussion around it. Addressing diversity in input modalities has become a challenge the research team has devoted significant time to but has not yet overcome. Voice control was part of the initial design plan for the ALS-SimVR due to an indication from other health care settings [46]; however, implementing it is not without significant challenges. In addition to the computational and algorithmic challenges, there is the issue of a hugely diverse cohort of individuals, all of whom often use different language to convey the same information using expressions drawn from a diverse and specific lexicon [47].

As a result, we propose other considerations for diversifying input modalities, such as the use of haptic inputs or hand gestures. The use of simple Oculus Go hand controllers in this study was met with mixed opinions by the participants due to the limited number of actions supported by the controller. The link between the choice of technology (or controllers) and simulation experience in medical training is not trivial, and yet lacking in the literature. Although our participants indicated a preference for diverse modalities that were not supported through the simple interactions offered by Oculus Go, our observation reveals the device did not interfere greatly with the participant's ability to perform the required tasks within ALS-SimVR. One explanation for this could be that whilst there were some limitations in the design of the ALS-SimVR app, many of the actions and interactions required were still achievable. However, the heuristics usability testing is designed to identify "issues" rather than highlight the positives, which may have led to this emphasis on the lack of diverse modalities. Further investigation into this matter is therefore warranted in future research.

Diversifying input modalities can potentially impact the experience of agency [48]. Agency in VR is generally defined as the experience of controlling actions and their consequences in the virtual environment [48]. Previous research has illustrated that the sense of agency in VR is diminished when relying on voice commands instead of hand controllers [48]. Therefore, the challenge with using voice in VR apps similar to ALS-SimVR is to combine it with other modalities that are meaningful to users and give them a sense of control. Our exploratory study did not address this issue. Further research is needed to identify which design features can improve the sense of agency in VR clinical training apps.

We found that assessing the experience of agency is particularly challenging. Agency is not directly addressed through the questionnaires we used in our research, but there are a number of heuristic items that may be linked to the experience of agency in the context of ALS-SimVR. For instance, support for learning, ability to coordinate actions, and compatibility of the app with the user's task and domain. These items align with the definition of agency as the ability to control actions. Additionally, the results from the PQ indicated the satisfaction of our participants with the "possibility to act", which can be interpreted as closely linked to agency. We propose that future research should specifically examine effective ways of evaluating agency in interactive VR experiences in medical training.

Limitations

A number of design and methodological limitations in our study warrant further research. Enabling clinicians to change their mind or edit decisions in ALS-SimVR posed challenges for the designers. As soon as any user choice is made in ALS-SimVR, it is logged and recorded to the database to generate feedback during the game and compile the "result" report at the end. The database was not designed to be edited once a decision was made, which means less flexibility of actions for users. Creating an updatable database requires more powerful hardware processing capability and continuous reviewing. It would also require a lot more coding development to allow the interruption of ongoing animations triggered by previous decisions. These were outside the available resources for the initial ALS-SimVR prototype used in this study.

From the methodological perspective, we intended to conduct an in-depth study; however, our sample size is small and from a clinically nondiverse background and included participants with little to no VR experience. As such, we received comments regarding application orientation that may not be applicable to more experienced users. Lack of VR experience made the learning curve of the app more challenging for our participants, as they had to familiarize themselves with a new technology. In addition, the prototype did not have an in-built orientation tutorial and the participants were given a brief verbal orientation. Additionally, with the high level of clinical expertise of the study group, there is a potential bias from the high level of clinical performance within the application. Future research should address these limitations. The findings from our qualitative study should be further validated through more testing and quantitative methods should be used to objectively address the five proposed themes.

ALS-SimVR: Next Steps

Previous research had suggested that high-fidelity interactive training simulations can foster clinician confidence [49].

However, in our study, new questions emerged on the effects of fidelity and realism within the immersive virtual clinical training environments. Questions targeting the realism of ALS-SimVR concern participant engagement and reflection on lived real-world experiences and underline the need for more immersive visuals and audio. The orderly and “sterile” nature of the current scenario detracted from the overall realism of the experience for users as it did not align with the common experience of clinicians responding to these events. This perceived sterility also leads us to believe there is further scope to explore the role of mental models and advanced roles within the ALS-SimVR.

The responsibility of being the team leader in an ALS scenario requires establishing and implementing mental models with a high degree of awareness toward potential consequences. For the effective management of a cardiac arrest or significant clinical deterioration, a team leader must be able to preplan their actions [50]. The ALS-SimVR app did not provide the opportunity for clinicians to demonstrate this preplanning. The linear nature of the decision-making process within the app and inability to stack or preset actions throughout the scenario distracted some users. A future iteration of ALS-SimVR will need to support this type of processing.

ALS is an advanced skill set requiring advanced roles and interventions [51]. The findings of this study highlighted that the challenge of providing users with the ability to perform these interventions must be considered and ideally overcome to maintain engagement. Our participants frequently brought up the importance of practicing advanced roles during interviews and this will inform future iterations of ALS-SimVR.

Since the completion of this study, a number of these identified issues have been addressed and included in subsequent versions of the application [52]. The application now runs on the Oculus Quest headset, which allows for greater processing power, higher resolution, and six degrees of freedom [53]. ALS-SimVR now has randomization of events and results to allow for greater replay ability, user performance is recorded and available for their own viewing and reflection on a web portal, and a tutorial has been added to standardize the introduction to the application. Finally, the animations and visualizations have also been improved.

Moving forward, the research team will gather additional data using other iterations of the application to further guide design and development for this application and VR-based educational design in general.

Conclusion

Interactive VR simulations and gameplays are increasingly used to augment traditional methods of clinical training such as ALS. However, there is limited information about how design can support the user experience and engagement needs in such VR apps. This study presents user-centered research on ALS-SimVR, an application designed to deliver ALS training. Based on interviews and usability testing of ALS-SimVR with 10 experienced clinicians, we identified five main areas of design considerations which should be considered when developing VR applications in this context: affordances, agency, diverse input modalities, mental models, and advanced roles. It was determined that a desirable user experience that supports ALS training and learning needs in VR must afford autonomous interactions within the VR app to allow users to complete the required clinical actions and make decisions in a way that is familiar to them and replicates real life. Diversifying input modalities to include voice commands within the VR environment may enhance the trainees’ experience of realism; however, this must be approached with care and combined with other modalities. Additionally, the tasks in VR must align with clinicians’ mental models for the role they are assigned in the game and how they would normally approach that challenge (eg, treating a patient). This may require careful consideration of interactions with other advanced roles in VR. The scale and spectrum of the interventions possible in this setting can provide challenges in creating a meaningful and realistic experience. The insights generated in this study can inform the design and testing of other similar VR-based training applications in health care, which is an area that requires further research [54].

Conflicts of Interest

None declared.

References

1. Castelao EF, Russo SG, Riethmüller M, Boos M. Effects of team coordination during cardiopulmonary resuscitation: a systematic review of the literature. J Crit Care 2013 Aug;28(4):504-521. [doi: 10.1016/j.jcrc.2013.01.005] [Medline: 23602030]
2. ANZCOR Adult Cardiorespiratory Arrest Flowchart. Australian Resuscitation Council. 2016. URL: https://resus.org.au/guidelines/flowcharts-3/ [accessed 2020-05-01]
3. Barsuk JH, Cohen ER, Wayne DB, Siddall VJ, McGaghie WC. Developing a Simulation-Based Mastery Learning Curriculum: Lessons From 11 Years of Advanced Cardiac Life Support. Simul Healthc 2016 Feb;11(1):52-59. [doi: 10.1097/SIH.0000000000000120] [Medline: 26536342]
4. McEvoy MD, Field LC, Moore HE, Smalley JC, Nietert PJ, Scarbrough SH. The effect of adherence to ACLS protocols on survival of event in the setting of in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):82-87 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.019] [Medline: 24103233]

5. Bingham AL, Sen S, Finn LA, Cawley MJ. Retention of advanced cardiac life support knowledge and skills following high-fidelity mannequin simulation training. Am J Pharm Educ 2015 Feb 17;79(1):12 [FREE Full text] [doi: 10.5688/ajpe79112] [Medline: 25741028]
6. Leigh G, Stueben F, Harrington D, Hetherman S. Making the Case for Simulation-Based Assessments to Overcome the Challenges in Evaluating Clinical Competency. Int J Nurs Educ Scholarsh 2016 May 13;13. [doi: 10.1515/ijnes-2015-0048] [Medline: 27176749]
7. Taplin J, McConigley R. Advanced life support (ALS) instructors experience of ALS education in Western Australia: a qualitative exploratory research study. Nurse Educ Today 2015 Apr;35(4):556-561. [doi: 10.1016/j.nedt.2014.12.017] [Medline: 25586006]
8. Flin R, Maran N. Identifying and training non-technical skills for teams in acute medicine. Qual Saf Health Care 2004 Oct;13 Suppl 1:80-84 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i80] [Medline: 15465960]
9. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 04;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
10. Cheng A, Nadkarni VM, Mancini MB, Hunt EA, Sinz EH, Merchant RM, et al. Resuscitation Education Science: Educational Strategies to Improve Outcomes From Cardiac Arrest: A Scientific Statement From the American Heart Association. Circulation 2018 Aug 07;138(6):82-122. [doi: 10.1161/cir.0000000000000583]
11. Chen J, Bellomo R, Flabouris A, Hillman K, Finfer S. The relationship between early emergency team calls and serious adverse events*. Critical Care Medicine 2009;37(1):148-153. [doi: 10.1097/ccm.0b013e3181928ce3]
12. Chen J, Ou L, Flabouris A, Hillman K, Bellomo R, Parr M. Impact of a standardized rapid response system on outcomes in a large healthcare jurisdiction. Resuscitation 2016 Oct;107:47-56. [doi: 10.1016/j.resuscitation.2016.07.240] [Medline: 27507434]
13. Pain C, Green M, Duff C, Hyland D, Pantle A, Fitzpatrick K, et al. Between the flags: implementing a safety-net system at scale to recognise and manage deteriorating patients in the New South Wales Public Health System. Int J Qual Health Care 2017 Feb 01;29(1):130-136. [doi: 10.1093/intqhc/mzw132] [Medline: 27920243]
14. Sachedina AK, Blissett S, Remtulla A, Sridhar K, Morrison D. Preparing the Next Generation of Code Blue Leaders Through Simulation. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare 2019;14(2):77-81. [doi: 10.1097/sih.0000000000000343]
15. Nickel F, Brzoska JA, Gondan M, Rangnick HM, Chu J, Kenngott HG, et al. Virtual Reality Training Versus Blended Learning of Laparoscopic Cholecystectomy. Medicine 2015;94(20):764. [doi: 10.1097/md.0000000000000764]
16. Aksoy E. Comparing the Effects on Learning Outcomes of Tablet-Based and Virtual Reality-Based Serious Gaming Modules for Basic Life Support Training: Randomized Trial. JMIR Serious Games 2019 May 01;7(2) [FREE Full text] [doi: 10.2196/13442] [Medline: 31042153]
17. Ahmadpour N, Keep M, Janssen A, Rouf AS, Marthick M. Design Strategies for Virtual Reality Interventions for Managing Pain and Anxiety in Children and Adolescents: Scoping Review. JMIR Serious Games 2020 Jan 31;8(1):e14565 [FREE Full text] [doi: 10.2196/14565] [Medline: 32012042]
18. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
19. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual Reality Clinical Research: Promises and Challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
20. Chen CJ. Theoretical bases for using virtual reality in education. Themes in Science and Technology Education 2010;2(1-2):71-90.
21. Saxena N, Kyaw BM, Vseteckova J, Dev P, Pradeep P, Lim KT, et al. Virtual reality environments for health professional education. Cochrane Library 2018. [doi: 10.1002/14651858.CD012090.pub2]
22. Ryan ML. Immersion vs. Interactivity: Virtual Reality and Literary Theory. SubStance 1999;28(2):110-137. [doi: 10.2307/3685793]
23. In Hospital ACLS Virtual Reality Training. Health Scholars. 2020 May 29. URL: https://healthscholars.com/acls/ [accessed 2020-05-29]
24. VR for healthcare. dualgood. 2020 May 29. URL: https://www.dualgoodhealth.com/healthcare [accessed 2020-05-29]
25. Panchal AR, Berg KM, Kudenchuk PJ, Del Rios M, Hirsch KG, Link MS, et al. 2018 American Heart Association Focused Update on Advanced Cardiovascular Life Support Use of Antiarrhythmic Drugs During and Immediately After Cardiac Arrest: An Update to the American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2018 Dec 04;138(23). [doi: 10.1161/cir.0000000000000613]
26. ANZCOR Adult Cardiorespiratory Arrest Flowchart. Australian Resuscitation Council. 2016 Jan 01. URL: https://resus.org.au/guidelines/flowcharts-3/ [accessed 2020-05-01]
27. Ijaz K, Ahmadpour N, Naismith SL, Calvo RA. An Immersive Virtual Reality Platform for Assessing Spatial Navigation Memory in Predementia Screening: Feasibility and Usability Study. JMIR Ment Health 2019 Sep 03;6(9):e13887 [FREE Full text] [doi: 10.2196/13887] [Medline: 31482851]

28. Binks A, LeClair R. Putting theory into practice: a method for generating useful pre-class materials to enhance student engagement. MedEdPublish 2018 Jul 20;7(3). [doi: 10.15694/mep.2018.0000148.1]
29. Krokos E, Plaisant C, Varshney A. Virtual memory palaces: immersion aids recall. Virtual Reality 2018 May 16;23(1):1-15. [doi: 10.1007/s10055-018-0346-3]
30. Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, et al. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science 2013 Nov 29;342(6162):1111-1114 [FREE Full text] [doi: 10.1126/science.1244056] [Medline: 24288336]
31. Vankipuram A, Khanal P, Ashby A, Vankipuram M, Gupta A, DrummGurnee D, et al. Design and Development of a Virtual Reality Simulator for Advanced Cardiac Life Support Training. IEEE J. Biomed. Health Inform 2014 Jul;18(4):1478-1484. [doi: 10.1109/jbhi.2013.2285102]
32. Walz SP, Deterding S. The gameful world: Approaches, issues, applications. Massachusetts: The MIT Press; 2015.
33. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane database of systematic reviews 2009 Jan 21. [doi: 10.1002/14651858.cd006575]
34. Smith SJ, Farra S, Ulrich DL, Hodgson E, Nicely S, Matcham W. Learning and Retention Using Virtual Reality in a Decontamination Simulation. Nursing Education Perspectives 2016;37(4):210-214. [doi: 10.1097/01.nep.0000000000000035]
35. Frengley RW, Weller JM, Torrie J, Dzendrowskyj P, Yee B, Paul AM, et al. The effect of a simulation-based training intervention on the performance of established critical care unit teams. Critical Care Medicine 2011;39(12):2605-2611. [doi: 10.1097/ccm.0b013e3182282a98]
36. Alvarez J, Djaouti D. An introduction to serious game definitions and concepts. In: Serious games & simulation for risks management. 2011 Presented at: Proceedings of the serious games and simulation for risks management workshop; December 5 2011; Observatoire de Paris - France p. 11-15.
37. Moore N, Yoo S, Ahmadpour N, Tommy R, Brown M, Poronnik P. ALS-SimVR: Advanced Life Support Virtual Reality Training Application. In: ACM Digital Library. 2019 Nov Presented at: 25th ACM Symposium on Virtual Reality Software and Technology; 2019; Parramatta p. 1-2. [doi: 10.1145/3359996.3365051]
38. Maskeliūnas R, Damaševičius R, Blažauskas T, Paulauska A, Paulauskas L, Chiatti C. IDO: modelling a serious educational game based on hands on approach for training dementia carers. IJET 2018 May 16;7(2.28):143-146. [doi: 10.14419/ijet.v7i2.28.12898]
39. Sutcliffe A, Gault B. Heuristic evaluation of virtual reality applications. Interacting with Computers 2004 Aug;16(4):831-849. [doi: 10.1016/j.intcom.2004.05.001]
40. Robillard G, Bouchard S, Cournoyer LG. Validation canadienne-française de deux mesures importantes en réalité virtuelle: Immersive Tendencies Questionnaire et le Presence Questionnaire. In: Poster presented at the 25e congrès annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières. 2002 Nov 01 Presented at: 25e congrès annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP); November 1; Trois-Rivières.
41. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
42. Evans SK, Pearce KE, Vitak J, Treem JW. Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research. J Comput-Mediat Comm 2016 Dec 26;22(1):35-52. [doi: 10.1111/jcc4.12180]
43. Kononowicz AA, Berman AH, Stathakarou N, McGrath C, Bartyński T, Nowakowski P, et al. Virtual Patients in a Behavioral Medicine Massive Open Online Course (MOOC): A Case-Based Analysis of Technical Capacity and User Navigation Pathways. JMIR Med Educ 2015 Sep 10;1(2):e8 [FREE Full text] [doi: 10.2196/mededu.4394] [Medline: 27731844]
44. Yasui K, Ryuta K, Tshiya A. Immersive virtual reality supporting content for evaluating interface using oculus rift and leap motion. International Journal of Innovative Computing, Information and Control 2019 Apr;15(2):591-605. [doi: 10.12792/icisip2018.033]
45. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. J Med Internet Res 2014 Dec 10;16(12):e251 [FREE Full text] [doi: 10.2196/jmir.3240] [Medline: 25498126]
46. Steinberg DM, Levine EL, Lane I, Askew S, Foley PB, Puleo E, et al. Adherence to self-monitoring via interactive voice response technology in an eHealth intervention targeting weight gain prevention among Black women: randomized controlled trial. J Med Internet Res 2014 Apr 29;16(4):e114 [FREE Full text] [doi: 10.2196/jmir.2996] [Medline: 24780934]
47. Chen J, Druhl E, Polepalli Ramesh B, Houston TK, Brandt CA, Zulman DM, et al. A Natural Language Processing System That Links Medical Terms in Electronic Health Record Notes to Lay Definitions: System Development Using Physician Reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]
48. Winkler P, Stiens P, Rauh N, Franke T, Krems J. How latency, action modality and display modality influence the sense of agency: a virtual reality study. Virtual Reality 2019 Sep 21. [doi: 10.1007/s10055-019-00403-y]
49. Blum CA, Borglund S, Parcells D. High-fidelity nursing simulation: Impact on student self-confidence and clinical competence. International Journal of Nursing Education Scholarship 2010 Jun 03;7(1). [doi: 10.2202/1548-923x.2035]
50. Andersen PO, Jensen MK, Lippert A, Østergaard D. Identifying non-technical skills and barriers for improvement of teamwork in cardiac arrest teams. Resuscitation 2010 Jun;81(6):695-702. [doi: 10.1016/j.resuscitation.2010.01.024] [Medline: 20304547]

51. Yang CW, Yen ZS, McGowan JE, Chen HC, Chiang WC, Mancini ME, et al. A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. Resuscitation 2012 Sep;83(9):1055-1060. [doi: 10.1016/j.resuscitation.2012.02.027] [Medline: 22391016]
52. 2020. ALS-SimVR. 2020 Jul 26. URL: https://frle.webflow.io/als-sim-vr/als-research [accessed 2020-07-26]
53. Facebook. 2020. Oculus Quest.: Facebook; 2020 Jan 01. URL: https://www.oculus.com/quest/?locale=en_US [accessed 2020-05-01]
54. Taveira-Gomes T, Ferreira P, Taveira-Gomes I, Severo M, Ferreira MA. What Are We Looking for in Computer-Based Learning Interventions in Medical Education? A Systematic Review. J Med Internet Res 2016 Aug 01;18(8):e204 [FREE Full text] [doi: 10.2196/jmir.5461] [Medline: 27480053]

Abbreviations

ACLS: Advanced Cardiac Life Support
ALS: Advanced Life Support
ALS-SimVR: Advanced Life Support Simulation in VR
HMD: head-mounted devices
PQ: Presence Questionnaire
VE: virtual environments
VR: virtual reality

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by M Eckert, K Glover, B Liu, S Gallagher; comments to author 18.06.20; revised version received 25.06.20; accepted 26.07.20; published 07.08.20.

Please cite as:
Moore N, Yoo S, Poronnik P, Brown M, Ahmadpour N
Exploring User Needs in the Development of a Virtual Reality–Based Advanced Life Support Training Platform: Exploratory Usability Study
JMIR Serious Games 2020;8(3):e20797
URL: http://games.jmir.org/2020/3/e20797/
doi:10.2196/20797
PMID:32763877

Original Paper

A Graded Exposure, Locomotion-Enabled Virtual Reality App During Walking and Reaching for Individuals With Chronic Low Back Pain: Cohort Gaming Design

Rebecca White Hennessy[1], BS; Deanna Rumble[2], PhD; Mike Christian[3], BS; David A Brown[4], PT, PhD; Zina Trost[5], PhD

[1]PhD Program in Rehabilitation Science, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States
[2]Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, United States
[3]From The Future LLC, Denton, TX, United States
[4]School of Health Professions, University of Texas Medical Branch, Galveston, TX, United States
[5]Department of Physical Medicine and Rehabilitation, Virginia Commonwealth University, Richmond, VA, United States

Corresponding Author:
Rebecca White Hennessy, BS
PhD Program in Rehabilitation Science
School of Health Professions
University of Alabama at Birmingham
1716 9th Avenue South
Birmingham, AL
United States
Phone: 1 205 975 3592
Email: rwhite31@uab.edu

Abstract

Background: Chronic low back pain (cLBP) can interfere with daily activities, and individuals with elevated pain-related fear (also known as kinesiophobia or the fear of injury due to movement) can develop worse long-term disability. Graded exposure (GEXP) protocols use successive participation in avoided activities to help individuals overcome fearful movement appraisals and encourage activity. We sought to develop a series of GEXP virtual reality (VR) walking and reaching scenarios to increase the exposure and engagement of people with high kinesiophobia and cLBP.

Objective: This study aims to (1) determine GEXP content validity of the VR application and (2) determine the feasibility of individuals with cLBP performing locomotion-enabled physical activities.

Methods: We recruited 13 individuals with cLBP and high pain-related fear to experience six VR modules, which provide progressive movement exposure over three sessions in a 1 week period. At session 1, participants ranked each module by likelihood to avoid and assigned an expected pain and concern for harming their back rating to each module. Participants provided a rating of perceived exertion (RPE) after experiencing each module. To test feasibility, we administered the system usability scale (SUS) and treatment evaluation inventory (TEI) following the final session. In addition, we measured pain and pain-related fear at baseline and follow-up.

Results: The 12 participants who completed the study period assigned higher avoidance (P=.002), expected pain (P=.002), and expected concern (P=.002) for session 3 modules compared with session 1 modules. RPE significantly increased from session 1 (mean 14.8, SD 2.3) to session 3 (mean 16.8, SD 2.2; P=.009). The VR application showed positive feasibility for individuals with cLBP through acceptable SUS (mean 76.7, SD 13.0) and TEI (mean 32.5, SD 4.9) scores. Neither pain (P=.20) nor pain-related fear (P=.58) changed significantly across sessions.

Conclusions: The GEXP VR modules provided progressive exposure to physical challenges, and participants found the VR application acceptable and usable as a potential treatment option. Furthermore, the lack of significant change for pain and pain-related fear reflects that participants were able to complete the modules safely.

(JMIR Serious Games 2020;8(3):e17799) doi:10.2196/17799

KEYWORDS

virtual reality; chronic low back pain; walking; rehabilitation; virtual reality exposure therapy

Introduction

Background

Chronic low back pain (cLBP)—low back pain present for longer than 3 months—is a common symptom with an estimated lifetime prevalence of 80% and is the second leading cause of disability in the United States [1,2]. Treatment guidelines include implementing self-care strategies (remain active, apply superficial heat, etc) and pharmacological (acetaminophen, nonsteroidal anti-inflammatory drugs, etc) and nonpharmacological treatments (exercise therapy, cognitive behavioral therapy, etc), but many cases resist traditional treatment [3]. Historically, the biomedical model of pain described the experience of pain solely through biological mechanisms and suggested a predictable and linear relationship between pain and tissue damage. However, this straightforward relationship has failed to explain many clinical observations of pain and has led to the uptake of a more biopsychosocial approach to chronic pain, which considers the experience of pain as a dynamic interaction between biological, sociocultural, and psychological factors [4,5].

The fear-avoidance model (FAM), a widely used theory that attempts to explain the development of chronic disability after a back injury, identifies pain-related fear as a central cognition that predicts long-term disability after musculoskeletal injury [6-9]. Elevated pain-related fear is the belief that pain always signals serious tissue damage, and the FAM postulates that individuals with high pain-related fear will avoid physical activities that they believe will further exacerbate pain, trigger reinjury, or prevent recovery. Elevated pain-related fear greatly affects movement quality and is related to slower walking speeds [10], modified reaching strategies [11], and reduced lifting ability [6]. In line with the FAM, an estimated 7.6 million Americans report disability related to back problems, including cLBP [2]. Accordingly, interventions that aim to increase physical function and decrease disability may have greater personal and societal impact than interventions that solely focus on decreasing pain intensity [12-14].

Graded Exposure as an Approach to Alleviate Pain-Related Fear

Treatments based on the FAM employ strategies to reduce pain-related fear and activity avoidance. Graded exposure (GEXP) is one such treatment, in which individuals with cLBP and high pain-related fear rank activities based on fearfulness and then progressively confront these fearful appraisals through activity [15]. The photographic series of daily activities (PHODA) is a measurement tool commonly used to capture how individuals rank expected pain and the harm of different daily activities [16,17]. The PHODA includes activities that participants can rank by perceived harm, such as walking while carrying shopping bags, twisting to take books off a shelf, and mowing the lawn.

GEXP protocols aim to correct catastrophic misinterpretations of pain sensations and reduce expectations of harm to the back, leading to functional improvements. A small number of randomized controlled trials have demonstrated that GEXP may clinically reduce fear of movement and disability among cLBP patients [18,19]. However, low patient engagement, high dropout, and limited ability to provide exposure to specific activities may limit the success of GEXP. Additionally, GEXP in a clinical setting relies on the availability of physical space and quality of clinical props to model relevant and lifelike exposure scenarios, which often results in training scenarios that are uninspiring and irrelevant to patient goals [20].

Virtual Reality as a Therapeutic Tool to Deliver GEXP

Virtual reality (VR) is a tool that can generate environments not otherwise possible in a clinical or laboratory setting. Previous research suggests that virtual environments can enhance rehabilitative training and improve physical outcomes [21-24]. GEXP VR protocols may enable clinicians to prescribe training scenarios that are not feasible in a traditional clinical setting. Specifically, VR may enhance traditional GEXP by offering tailored training based on patient goals, reduced clinician workload, and improved patient monitoring through movement tracking [25]. VR can generate an unlimited number of objects and environments that may enhance patient interaction and improve the intrinsic motivation of GEXP therapy.

Researchers and clinicians can categorize movement tasks by body orientation (providing stability or transport), environment (stationary or in motion), and object manipulation requirements (holding an object or not) [26,27]. Several VR applications have already shown promising feasibility and usability in individuals with cLBP [28,29], but to date, they have only provided exposure to back-challenging activities (ie, reaching) from a stationary position. Stationary reaching tasks can include object manipulation (eg, reaching with an object) and a dynamic environment, but they limit users to reaching from a seated or standing position. Although this may be appropriate for some users, this approach neglects reaching activities that are more relevant to daily living that require body transport. Thus, these systems limit users and force training to stop before more complex tasks, such as movement that requires walking.

In response, we have developed an engaging, locomotion-enabled GEXP VR application to address the lack of applications that provide progressive movement challenges for individuals with cLBP. The VR application, Lucid, consists of six 3-min modules that challenge participants to complete engaging activities in VR that require progressively more challenging walking and reaching movements in real life. The progressive modules deliver exposure over 3 study sessions in a weeklong training period, where users experience 2 modules at each session.

Objectives

Before evaluating the efficacy of the GEXP VR application on pain-related health outcomes, we needed to evaluate the basic parameters of the application's GEXP content and establish how potential users respond to the application. In this study, we aimed to measure the *GEXP content* (ie, How well did the module goals gradually challenge participants with fear-inducing movements) and *feasibility* (ie, Is it possible for participants with cLBP to use the app?). To measure feasibility, we specifically examined *usability* (ie, Can users accomplish module goals with effectiveness, efficiency, and satisfaction? [30]), *safety* (ie, Can users complete the module tasks without experiencing harm?), and *acceptance* (ie, Do users find the potential treatment option fair, reasonable, and appropriate? [31]).

Hypothesis 1: GEXP Content

Before exposure to the VR modules, participants would report higher expected ranked avoidance, higher expected pain ratings, and higher expected back-related concern ratings for the more challenging modules compared with less challenging modules.

Hypothesis 2: Acceptability and Usability

Participants would report that the VR application is acceptable and usable, as determined by acceptability and usability scores above the respective cutoff values.

Hypothesis 3: Safety

The VR application would be safe for participants with cLBP to complete, as determined by no significant increase in pain or pain-related fear over the study period.

Methods

Equipment

We used a commercially available self-driven treadmill and a VR system to deliver the VR walking experience (Figure 1). Participants wore an HTC-Vive head-mounted display (HMD). Depending on the module, participants held either a regular HTC-Vive controller in each hand or a heavier, custom-made controller weighing approximately 2 lbs (Figure 2). The KineAssist-MX (KA-MX), a specialized self-driven treadmill, allowed participants to set their own walking pace in real time and naturally translate in the VR environment [32,33]. The KA-MX provides a safe walking environment, and several other studies involving both nonimpaired and clinical populations have used it [34-37]. The KA-MX consists of a pelvic mechanism that attaches to participants via a pelvic harness and allows participants normal hip range of motion [38]. As a participant intends to take a step, the pelvic mechanism detects the force generated by the participant's movement, and a built-in algorithm converts this force to a proportional speed on the treadmill belt. Participants could speed up, speed down, or change direction simply by moving with the intended force in the intended direction. We added a 2-kg upward force through the device's unweighing function to counteract the weight of the pelvic mechanism on the participants. The KA-MX sent the treadmill speed to a computer via an ethernet cable, and the HTC-Vive system tracked head and hand position. Users walked, reached, and crouched in real life to move around in the VR environment.

Figure 1. Walking and reaching virtual reality application setup.

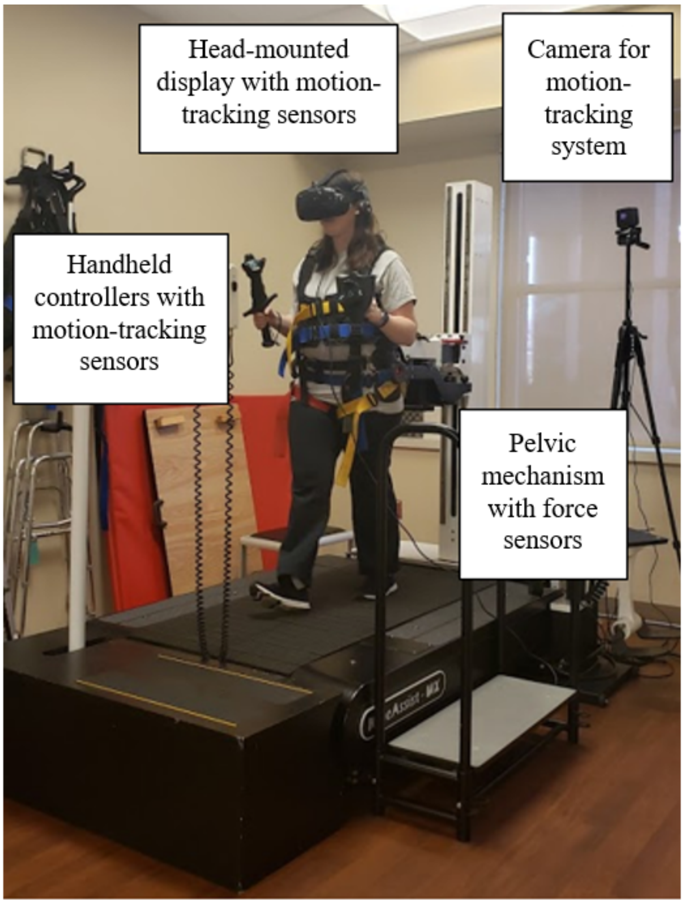


Figure 2. Weighted sword and shield hand-held controllers.

VR Development

We designed and developed the novel VR application, Lucid, that encourages participants to practice real-world movement tasks in a fun and engaging VR world. First, we identified functional movements used in real-world activities by using the movements included in the PHODA. The PHODA includes 8 movement types (ie, lifting, bending, turning, reaching, falling, intermittent load, unexpected movement, long-lasting load instance, or sit with limited dynamics) performed in either static or dynamic positions [39]. From these movements, we selected reaching, bending, and long-lasting loads to incorporate into the VR walking challenges because they have been used in previous studies to provide a graded challenge for individuals with cLBP [16,40]. Second, we designed VR activities that required different levels of each movement task to complete. These VR activities included goals such as fighting monsters, crouching under branches, saving animals that challenged participants to reach, bend to get under obstacles, and walk quickly. We then incorporated the VR activities into six 3-min modules that encouraged participants to complete progressively more difficult combinations of the movement tasks (Table 1). We created the VR application, named Lucid, specifically for this study through a Small Business Innovation Research grant partnership with From the Future, LLC. We used Unity to develop the VR application, which ran on a Windows 10–based personal computer.

Table 1. Movement requirements and activity goals for the virtual reality modules.

Session	Module	Movement requirements	VR[a] activity goal
1 (low-intensity challenge)	1	• Walking: any pace • Reaching: requires one hand • Bending: no • Carry weights: no	"Walk at your own pace and rid the realm of monsters. Swing your sword to damage foes and block their attacks with your shield."
1 (low-intensity challenge)	2	• Walking: walking quickly • Reaching: requires one hand • Bending: no • Carry weights: no	"Walk at an increased pace to save as many animals as you can. Monsters have started to prey on the wildlife, and it's up to you to save the animals before the monster consumes them."
2 (medium-intensity challenge)	3	• Walking: any pace • Reaching: requires both hands • Bending: no • Carry weights: no	"The monsters have desolated the land, and it's up to you to collect food and coins for the realm. You are given two swords to reach both your foes and your items in all directions."
2 (medium-intensity challenge)	4	• Walking: any pace • Reaching: requires one hand • Bending: yes • Carry weights: no	"Crouch under trees and tunnels to explore more of the realm. You'll want to make sure you avoid limbs and the ceiling, or you'll bring your journey to an end."
3 (high-intensity challenge)	5	• Walking: any pace • Reaching: requires both hands • Bending: yes • Carry weights: yes	"Wield a weighted sword and shield while you crouch under trees and tunnels to explore more of the realm. You'll want to make sure you avoid limbs and the ceiling, or you'll bring your journey to an end."
3 (high-intensity challenge)	6	• Walking: walking quickly • Reaching: requires both hands • Bending: yes • Carry weights: yes	"Wield a weighted sword and shield to defeat your enemies."

[a]VR: virtual reality.

In the HMD, participants navigated a walking path and utilized a VR sword and shield (Figure 3). Participants walked to translate along the path and moved the HTC-Vive controllers to wield the sword and shield. The VR application presented module challenges throughout the trail. Participants did not have to reach a certain distance in the trail or score a certain number of points, so not accomplishing the goal because of walking speed, reaching ability, or bending ability was not a possible outcome. Rather, we simply instructed the participants to do their best and focus on the module objective for the entire 3 min. Additionally, certain in-game collectible objects temporarily equipped the participants with a special ability (eg, longer sword) to help them with their goal.

Figure 3. View through the head-mounted display as a user combats a monster while holding the virtual reality sword and shield.

Participants

We recruited potential participants through fliers, through the web-based clinical trial registration, and from individuals who participated in previous studies in our lab. We included potential participants aged 18-65 years so that we could test the feasibility in a wide age range of individuals who may benefit from a GEXP intervention. We screened individuals over the phone, and individuals were deemed eligible if they self-reported low back pain for longer than 3 months, experienced interference caused by their back pain in daily life, and reported elevated pain-related fear. Participants were determined to have a high level of fear if they scored greater than 10 points on a 4-question pain-related fear screen that consisted of the 4 highest loaded items from the Tampa scale for kinesiophobia (TSK) [41]. Exclusion criteria included not passing a medical screen (eg, inability to stand for 15 min, recent fall), pregnancy, and significant medical conditions that impair movement ability (eg, arthritis and plantar fasciitis). Additionally, we excluded participants involved in an active training study or a legal claim so that our protocol did not influence other proceedings where pain status is an important outcome. Author DB, a physical therapist, interviewed each participant at the beginning of the study period to determine that each participant had low back pain in a chronic state and did not have a secondary condition that would interfere with participation.

There were 12 participants (aged 43-60 years) with cLBP and high pain-related fear that completed the study protocol (Table 2). One additional participant enrolled in the study but was unable to schedule the VR sessions within 1 week, and we excluded this participant from the analysis. Participants reported an average pain of 6.4 (SD 2.1) for the previous 7 days before the first study session on a numeric rating scale (0 for no pain to 10 for worst imaginable pain). In total, 58% (7/12) of participants were taking pain medications, none of the participants had undergone surgery, and none of the participants were in a physical therapy program at the time of the study. Most participants (8/12, 67%) had back pain for more than 5 years.

Table 2. Participant characteristics (N=12).

Characteristics	Values
Age (years), mean (SD)	54.3 (5.1)
Gender, n (%)	
Male	4 (33)
Female	8 (67)
Race, n (%)	
Black	12 (100)
MPQ-sf[a] PRI[b] (possible score range 0-45), mean (SD)	14.8 (9.4)
MPQ-sf VAS[c] (possible score range 0-100), mean (SD)	52.3 (23.6)
MPQ-sf PPI [d] , n (%)	
No pain	0 (0)
Mild	1 (8)
Discomforting	0 (0)
Distressing	10 (83)
Horrible	1 (8)
Excruciating	0 (0)
TSK[e] (possible score range 17-68), mean (SD)	44.2 (8.0)
Walking speed in VR[f] (m/s), mean (SD)	1.1 (0.2)

[a]MPQ-sf: McGill pain questionnaire-short form.
[b]PRI: pain rating index.
[c]VAS: visual analog scale.
[d]PPI: present pain intensity.
[e]TSK: Tampa scale for kinesiophobia (scores >37 are indicative of high fear).
[f]VR: virtual reality.

Study Design

Participants attended 3 VR sessions (sessions 1-3) over a 1-week period and a follow-up session 3 to 5 days after session 3. During each VR session, participants tested 2 Lucid VR modules in progressive order. Participants completed a baseline questionnaire at the beginning of session 1 and then a follow-up questionnaire at the follow-up session. The questionnaire at session 1 included basic demographics, including gender, age, and race. The institutional review board at the University of Alabama at Birmingham approved the study design, all participants provided written informed consent, and we compensated participants for their time. The study sessions were conducted at a laboratory in the University of Alabama at Birmingham.

Primary Measures of GEXP Content Validity (Hypothesis 1)

Our primary *GEXP content validity* outcome was the avoidance rank, expected pain, and expected concern participants assigned to each Lucid module at baseline. Developed in the same style as the PHODA, we developed a card for each module that showed an avatar performing the movement required in the module. We laid the 6 cards randomly in front of each participant. Participants organized the cards in a row from the activities they would be least to most likely to avoid. We discreetly recorded the card's position as the *ranked avoidance* (1 for least likely to avoid to 6 for most likely to avoid), which we used as a measure of how participants perceived the difficulty of the modules in relation to the other modules. Then, to encourage participants to more critically evaluate the tasks in each module, participants reported their *expected pain* if they were to perform the activity (0 for no pain to 100 for worst possible pain) and their *expected concern and worry for harming their back* (0 for not at all concerned to 10 for extremely concerned) for each card.

Secondary Measures of GEXP Content Validity

Secondarily, we also asked participants to provide their rating of perceived exertion (RPE) after each VR module to measure how participants rated their perceived effort during each VR module [42]. To administer the RPE, we asked participants, “How hard were you working during that activity?” after each module. The RPE scale ranges from 6 (no exertion) to 20 (maximal exertion), and we expected participants to assign higher exertion for the modules with higher challenge intensity. RPE is an accurate measure of perceived effort in individuals with cLBP [43]. Additionally, at session 3, we asked participants to rate the difficulty of the session modules compared with the previously experienced modules (questions included in

Multimedia Appendix 1). For each question, participants could respond from 0 (much easier) to 10 (much more difficult).

Primary Measures of Feasibility (Hypotheses 2 and 3)

At the follow-up session, we administered the treatment evaluation inventory-short form (TEI-sf) as the primary measure of *acceptability*. The TEI-sf is a 9-item questionnaire used to measure intervention acceptability from the participants' point of view. Scores range from 9 to 45, and scores above 27 are considered acceptable [31]. The TEI-sf is a valid measure [31], and several other studies have used it to measure acceptability in participants with back pain [44,45]. We also administered the system usability scale (SUS) at follow-up as our primary measure of *usability*. The SUS is a reliable 10-item usability measure and includes statements such as "I thought this system was easy to use" and "I felt confident using this system" [46,47]. Scores range from 0 to 100, and an above average score is 68 or higher. Our primary *safety* outcomes were the difference in pain and pain-related fear between baseline and follow-up, as measured by the McGill pain questionnaire-short form (MPQ-sf) [48] and the TSK [49], respectively. Both these constructs are important in chronic pain management, and we wanted to ensure that the GEXP VR application as a potential treatment option did not negatively influence them. The MPQ-sf is a valid measure that asks participants to respond to 15 pain descriptors (eg, throbbing and tender) on a 4-point Likert scale (0 for none to 3 for severe) to measure the person's pain rating index (PRI) [50]. The MPQ-sf also includes a visual analog scale (VAS) to rate current pain intensity on a continuous scale (0 for no pain to 100 for worst possible pain) and a present pain intensity index, which asks respondents to select 1 of 6 words to describe their pain (0 for no pain to 5 for excruciating pain). We administered the TSK at baseline and follow-up to measure pain-related fear. The TSK is a valid and reliable tool to measure fear of movement in individuals with cLBP [8,41,49]. TSK scores range from 17 to 68, with scores greater than 37 indicating elevated kinesiophobia [8].

Secondary Measures of Feasibility

To measure pain changes immediately following each module, we asked participants to mark their pain along a VAS before and after each module. VAS measurements are a valid and reliable method to measure pain in individuals with cLBP [51,52]. After each session, participants also filled out a custom-made questionnaire, which included statements such as "I felt challenged," "It was fun," and "I felt motivated" on a 0 (not at all) to 10 (extremely) scale (questions included in Multimedia Appendix 1).

Data Analysis

This study aimed to (1) determine the *GEXP content* validity of the VR application and (2) determine the *feasibility* of individuals with cLBP performing integrated physical activities. We descriptively summarized demographic data.

To examine the GEXP content validity of the graded modules, we used Friedman tests and Dunn post hoc pairwise comparisons with Bonferroni corrections to analyze ordinal data across sessions (avoidance rank and RPE). For continuous data (expected pain and expected concern), we performed a repeated measures analysis of variance (ANOVA) across the VR sessions and performed post hoc pairwise comparisons with Bonferroni corrections. We reported the means and SDs of how participants rated each session's difficulty compared with the previous sessions.

To evaluate the feasibility of the VR application, we calculated the means and SDs of the TEI-sf, SUS, and postsession questionnaire. We performed paired *t* tests to compare participant pain and pain-related fear from baseline to follow-up. We also used paired *t* tests to examine differences in VAS pain from before to after each module.

We performed Shapiro-Wilk normality tests to confirm the normality of the distribution for dependent variables. For all analyses, we collapsed the VR modules into 3 groups (session 1, session 2, and session 3), based on the 3 VR sessions. We set the alpha level of significance to .05 (two tailed) for all statistical tests. We checked the data for underlying assumptions, and data were described and analyzed using IBM SPSS 25 (IBM Corp).

Results

Primary Measures of GEXP Content Validity Results (Hypothesis 1)

Participants assigned higher avoidance, expected pain, and expected concern to the Lucid sessions that are designed to be more challenging (Table 3). A Friedman test revealed that the difference in avoidance was statistically significant (χ^2_2=15.1; *P*=.001). Dunn post hoc pairwise comparisons indicated that the difference between session 1 and session 3 was significant after Bonferroni correction (*P*=.002). The avoidance difference between sessions 1 and 2 and the avoidance difference between sessions 2 and 3 were not significant. A repeated measures ANOVA revealed significant differences across the sessions for expected pain ($F_{2,22}$=17.9; *P*<.001) and expected concern for harming their back ($F_{2,22}$=16.83; *P*<.001). Pairwise comparisons revealed that expected pain was significantly higher for session 3 than for session 1 (*P*=.002) and session 2 (*P*=.03) after Bonferroni correction. The difference in expected pain between session 1 and session 2 was also significant (*P*=.003). Similarly, pairwise comparisons revealed that expected concern was significantly higher for session 3 than for session 1 (*P*=.002) and session 2 (*P*=.008) after Bonferroni correction. The difference in expected concern between session 1 and session 2 was also significant (*P*=.03).

Table 3. Avoidance rank, expected pain, and expected concern for harm.

Measures and sessions	Values	95% CI
Avoidance rank (possible score range 0-6), median (IQR)		
Session 1	2.5 (1.4)[a]	2.0-2.8
Session 2	3.3 (1.8)	2.6-3.0
Session 3	5.5 (0.8)[a]	4.7-5.6
Expected pain (possible score range 0-100), mean (SD)		
Session 1	38.3 (25.4)[a,b]	22.2-54.5
Session 2	54.2 (25.2)[b,c]	38.2-70.2
Session 3	69.4 (22.2)[a,c]	55.3-83.5
Expected concern (possible score range 0-100), mean (SD)		
Session 1	42.3 (27.5)[b,c]	24.8-59.8
Session 2	57.5 (31.2)[b,c]	37.7-77.3
Session 3	74.4 (30.9)[a,c]	54.7-94.9

[a]Significant pairwise comparison between sessions 1 and 3.
[b]Significant pairwise comparison between sessions 1 and 2.
[c]Significant pairwise comparison between sessions 2 and 3.

Secondary Measures of GEXP Content Validity Results

The average RPE slightly increased over the sessions, and 83% (10/12) of the participants reported a higher exertion on session 3 compared with session 1 (Table 4). The Friedman test indicated that the difference in exertion between sessions was significant (χ^2_2=10.0; *P*=.007). Post hoc pairwise comparisons using Dunn method indicated that the difference between session 1 and session 3 was significant after Bonferroni correction (*P*=.009). The RPE difference between sessions 1 and 2 and the difference between sessions 2 and 3 were not significant.

Table 4. Rating of perceived exertion.

Sessions	Median (IQR)[a]	95% CI
1	14.0 (4.75)[b]	13.2-16.3
2	15.0 (4.75)	14.6-17.5
3	17.0 (4.75)[b]	15.4-18.2

[a]Possible scores from 6 to 20.
[b]Significant pairwise comparison between sessions 1 and 3.

On the difficulty rating questions, participants reported that the session 3 modules were more difficult than the session 1 modules (mean 6.1, SD 3.6) and session 2 modules (mean 5.9, SD 3.5).

Primary Measures of Feasibility Results (Hypotheses 2 and 3)

TEI: Acceptability

Participants responded positively to the VR application as an acceptable potential intervention for cLBP. The average TEI score was 32.5 (SD 4.9), which is above the acceptability cutoff score of 27. Scores ranged from 26 to 41, and 92% (11/12) of participants responded at the cutoff score or above.

SUS: Usability

The average SUS score was 76.7 (SD 13.0). Scores ranged from 52.5 to 92.5, and 75% (9/12) of the participants reported that the system was usable.

Pain and Pain-Related Fear: Safety

From baseline to follow-up, there were no overall changes in the MPQ-sf PRI (*P*=.20), MPQ-sf VAS (*P*=.73), or TSK (*P*=.58). In total, 67% (8/12) of the participants had improved MPQ-sf PRI scores at follow-up, and participants who improved showed an average decrease of 7.8 (SD 5.1) points on the PRI.

Secondary Measures of Feasibility Results

In most of the VR modules, there were no statistically significant changes in VAS pain ratings. However, VAS pain significantly increased during module 6 (*P*=.02), the most physically challenging module. In module 6, average post-VR pain rose to 51.5 (SD 32.01) from the 44.0 (SD 28.5) pre-VR pain levels. On the postsession questionnaire, participants responded positively to the modules (Multimedia Appendix 1). After session 3, participants responded that even though the modules were challenging (mean 9.0, SD 1.4), they were also motivating (mean 9.0, SD 1.3), enjoyable (mean 8.7, SD 1.8), and fun (mean 9.1, SD 1.3).

Discussion

Principal Findings

In this study, we aimed to (1) determine the GEXP content validity of the VR application and (2) determine the feasibility of individuals with cLBP performing integrated physical activities. Participant responses supported that the VR modules provided progressive exposure to fearfully perceived tasks by assigning greater avoidance to the modules designed to elicit greater fearful appraisals and present greater challenges. In concurrence with the FAM, participants assigned the highest expected pain and concern for harming their back to the modules that they ranked with higher avoidance. For feasibility, the TEI and SUS scores indicated that participants with high fear and cLBP found the VR application an acceptable approach to treat cLBP and usable as a system. Additionally, the VR application was safe, as participants successfully completed the GEXP VR protocol without negative effects on pain or pain-related fear.

The GEXP VR application is an important step in using VR as a potential treatment option because it allowed participants to gradually practice real-life movements through VR activities that require combinations of walking and reaching. Many daily tasks require a person to combine walking and reaching abilities, such as walking to open a door or carrying a bag of groceries. Previous applications have provided exposure to stationary tasks, but by incorporating these tasks into a walking environment, there may be greater potential to translate learning into real-world activities.

The participant feedback we captured is consistent with other studies that have tested VR apps designed for individuals with cLBP. Thomas et al [28] reported the feasibility and safety of a VR dodgeball intervention that encourages participants to use various amounts of lumbar flexion to achieve VR reaching and bending objectives. They reported that their participants were able to complete the VR activities with no adverse events and that their participants responded positively to the VR activity through agreeing with statements such as “The game was fun” and “The game encouraged me to move”. Yelvar et al [53] reported that passively viewing VR walking scenes during physical therapy may improve pain-related outcomes in individuals with cLBP. Additionally, Fowler et al [54] studied the use of VR to gradually expose veterans with chronic pain to progressively more involved movements. They found that gradual VR exposure was feasible but also reported that users rated the activities, designed to progressively deliver more exposure, with similar intensity ratings [54]. Although these applications are useful to train specific movement tasks and support that VR may be an acceptable way to address chronic pain, the VR application described in this study is the first app, to our knowledge, designed specifically to apply VR GEXP in an interactive walking protocol for individuals with cLBP.

Although we intentionally used a GEXP mechanism in the Lucid VR application to decrease the physical limitations caused by activity avoidance, other factors may have helped participants complete the modules. Distraction from pain is a well-studied mechanism commonly used in acute pain [20,55]. We added components in each module to increase movement exposure, such as adding coins to collect or animals to rescue. These novel components may have helped hold the attention of participants and maintained distraction. Many participants reported that they were less aware of their pain and that the modules distracted them from their pain, and this likely contributed to their expressed acceptance of the VR application. Therefore, although VR GEXP may be a targeted way to progressively challenge participants with cLBP, distraction may be an important component to include in future VR applications.

The lack of observed elevations for pain and pain-related fear across the study sessions reflects that individuals with high fear and cLBP were able to successfully complete the challenging activities without adverse consequences in these domains. In line with our long-term intervention goal to reduce deficits in physical ability caused by activity avoidance, participants with cLBP and elevated pain-related fear exhibited their ability to perform functional movement activities such as walking, reaching, carrying, and crouching despite experiencing pain and pain-related fear.

Our study goal was to test the GEXP content and feasibility of the VR application, which we designed to improve the physical abilities of individuals with cLBP by gradually exposing participants to more difficult challenges. Given that chronic pain can significantly interfere with one’s goals and ability to complete everyday activities, interventions that increase physical ability are valuable. This VR application allows individuals not only to interact in an interesting and challenging VR world, but it provided physical challenges that incorporated body transport and reaching movements in a dynamic and motivating environment.

Limitations

As our study design was to establish the content validity and feasibility of the GEXP app, all participants received the same activities in the same order. The lack of personalization could have limited the GEXP experience for some of our participants as we only provided exposure to these predefined movement combinations. We do not believe this limitation had a significant effect on our study outcomes as participants generally ranked the activities by avoidance in the expected order. In addition, although we used RPE to measure perceived effort, we did not ask participants to specifically rate the difficulty of each module given their back pain. Participants may have assigned greater effort to the higher intensity modules for reasons other than back-related challenges. Participants who walked faster may have had higher exposure as they were able to progress further along the trail in 3 min. Additionally, we only recruited 13 individuals to test the VR app, and this potentially limits our statistical power. To our knowledge, this is the first study to test a VR GEXP walking protocol for individuals with cLBP, and it was essential to determine whether individuals with cLBP could engage with the system and perform the challenges in the novel walking environment before expanding into a larger sample size.

Future Directions

The next step in this line of research is to explore the efficacy of the VR application on pain-related health outcomes. For this, we would need a larger sample size and a longer study duration. Although this study focused on GEXP content and feasibility, we would also need to explore how the VR exposure training translates to real-world activities by measuring changes in avoidance, disability, and physical function outside of the VR setting. Additionally, future iterations could improve the GEXP experience by allowing the participants to rank a greater number of activities and then experience modules tailored to how they rank their avoidance of each task. This would improve the personalization of the VR application and could allow participants to experience a tailored program more relevant to their daily living. Although our study focused on the general perceived effort of module activities, future studies could also specifically ask about the perceived back-related challenge of activities to better characterize the GEXP. In line with using VR to provide a tailored therapeutic experience, future studies should also explore how age, gender, and pain status influence participation in GEXP VR therapies.

Conclusions

We have established that the VR modules provided progressive challenges and were feasible for individuals with cLBP and high pain-related fear. The locomotion-enabled VR modules allowed users to freely walk and complete challenging physical activities in a motivating environment that participants thought was acceptable, usable, and safe. Expectation ratings, RPE, and module difficulty responses support that the sessions and comprising modules provided a progressive challenge, in line with GEXP protocols. Despite presenting activities likely for individuals with high fear to avoid, the graded VR walking challenges did not increase pain or fear of movement. The VR modules provided exposure to physical activity challenges that integrate reaching, walking, crouching, and carrying weights while also providing a safe bout of exercise and an enjoyable gaming experience.

Acknowledgments

This research was supported by the National Institutes of Health/National Institute of Child Health and Human Development of the National Institutes of Health under award number #1R43HD093522-01 and through a grant from the University of Alabama at Birmingham Health Services Foundation General Endowment Fund (PIs: Carmen Capo-Lugo [PT, PhD] and DB [PT, PhD]). The authors would like to thank Jennifer Uzochukwu and Caroline Springer for their assistance with data collection and data entry.

Authors' Contributions

ZT and DB designed the study and provided their expertise during game development so that the game elements incorporate important and challenging physical activities appropriate for individuals with chronic pain. MC designed, built, and integrated the VR application with the KA-MX. RH collected and analyzed the data, and DR collected data and assisted with the study design.

Conflicts of Interest

DB receives royalties on sales of the KineAssist MX device used in this study. MC with From the From the Future LLC sells virtual reality software as a KineAssist MX device add-on. All other authors have no financial interest to disclose.

Multimedia Appendix 1
Postsession questionnaire and responses.
[DOC File , 49 KB - games_v8i3e17799_app1.doc]

References

1. Rubin DI. Epidemiology and risk factors for spine pain. Neurol Clin 2007 May;25(2):353-371. [doi: 10.1016/j.ncl.2007.01.004] [Medline: 17445733]
2. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults--United States, 2005. MMWR Morb Mortal Wkly Rep 2009 May 1;58(16):421-426 [FREE Full text] [Medline: 19407734]
3. Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle P, Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 2;147(7):478-491. [doi: 10.7326/0003-4819-147-7-200710020-00006] [Medline: 17909209]
4. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 8;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]

5. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
6. Crombez G, Vlaeyen JW, Heuts PH, Lysens R. Pain-related fear is more disabling than pain itself: evidence on the role of pain-related fear in chronic back pain disability. Pain 1999 Mar;80(1-2):329-339. [doi: 10.1016/s0304-3959(98)00229-2] [Medline: 10204746]
7. Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, Vlaeyen JW. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. J Behav Med 2007 Mar;30(1):77-94. [doi: 10.1007/s10865-006-9085-0] [Medline: 17180640]
8. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 1995 Sep;62(3):363-372. [doi: 10.1016/0304-3959(94)00279-n] [Medline: 8657437]
9. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000 Apr;85(3):317-332. [doi: 10.1016/s0304-3959(99)00242-0] [Medline: 10781906]
10. Al-Obaidi SM, Al-Zoabi B, Al-Shuwaie N, Al-Zaabie N, Nelson RM. The influence of pain and pain-related fear and disability beliefs on walking velocity in chronic low back pain. Int J Rehabil Res 2003 Jun;26(2):101-108. [doi: 10.1097/00004356-200306000-00004] [Medline: 12799603]
11. Thomas JS, France CR. Pain-related fear is associated with avoidance of spinal motion during recovery from low back pain. Spine (Phila Pa 1976) 2007 Jul 15;32(16):E460-E466. [doi: 10.1097/BRS.0b013e3180bc1f7b] [Medline: 17632385]
12. Gardner T, Refshauge K, McAuley J, Goodall S, Hübscher M, Smith L. Patient led goal setting in chronic low back pain-what goals are important to the patient and are they aligned to what we measure? Patient Educ Couns 2015 Aug;98(8):1035-1038. [doi: 10.1016/j.pec.2015.04.012] [Medline: 25959985]
13. Staal JB, Rainville J, Fritz J, van Mechelen W, Pransky G. Physical exercise interventions to improve disability and return to work in low back pain: current insights and opportunities for improvement. J Occup Rehabil 2005 Dec;15(4):491-505. [doi: 10.1007/s10926-005-8030-3] [Medline: 16254751]
14. Sullivan MD, Ballantyne JC. Must we reduce pain intensity to treat chronic pain? Pain 2016 Jan;157(1):65-69. [doi: 10.1097/j.pain.0000000000000336] [Medline: 26307855]
15. George SZ, Wittmer VT, Fillingim RB, Robinson ME. Comparison of graded exercise and graded exposure clinical outcomes for patients with chronic low back pain. J Orthop Sports Phys Ther 2010 Nov;40(11):694-704 [FREE Full text] [doi: 10.2519/jospt.2010.3396] [Medline: 20972340]
16. Trost Z, France CR, Thomas JS. Examination of the photograph series of daily activities (PHODA) scale in chronic low back pain patients with high and low kinesiophobia. Pain 2009 Mar;141(3):276-282. [doi: 10.1016/j.pain.2008.11.016] [Medline: 19131166]
17. Leeuw M, Goossens ME, van Breukelen GJ, Boersma K, Vlaeyen JW. Measuring perceived harmfulness of physical activities in patients with chronic low back pain: the photograph series of daily activities--short electronic version. J Pain 2007 Nov;8(11):840-849. [doi: 10.1016/j.jpain.2007.05.013] [Medline: 17632038]
18. Leeuw M, Goossens ME, van Breukelen GJ, de Jong JR, Heuts PH, Smeets RJ, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. Pain 2008 Aug 15;138(1):192-207. [doi: 10.1016/j.pain.2007.12.009] [Medline: 18242858]
19. Woods MP, Asmundson GJ. Evaluating the efficacy of graded in vivo exposure for the treatment of fear in patients with chronic back pain: a randomized controlled clinical trial. Pain 2008 Jun;136(3):271-280. [doi: 10.1016/j.pain.2007.06.037] [Medline: 17716819]
20. Trost Z, Parsons TD. Beyond distraction: virtual reality graded exposure therapy as treatment for pain-related fear and disability in chronic pain. J Appl Biobehav Res 2014 Jun 10;19(2):106-126. [doi: 10.1111/jabr.12021]
21. Jaffe DL, Brown DA, Pierson-Carey CD, Buckley EL, Lew HL. Stepping over obstacles to improve walking in individuals with poststroke hemiplegia. J Rehabil Res Dev 2004 May;41(3A):283-292 [FREE Full text] [doi: 10.1682/jrrd.2004.03.0283] [Medline: 15543446]
22. Levac DE, Huber ME, Sternad D. Learning and transfer of complex motor skills in virtual reality: a perspective review. J Neuroeng Rehabil 2019 Oct 18;16(1):121 [FREE Full text] [doi: 10.1186/s12984-019-0587-8] [Medline: 31627755]
23. Holden MK. Virtual environments for motor rehabilitation: review. Cyberpsychol Behav 2005 Jun;8(3):187-211; discussion 212. [doi: 10.1089/cpb.2005.8.187] [Medline: 15971970]
24. Lozano-Quilis J, Gil-Gómez H, Gil-Gómez JA, Albiol-Pérez S, Palacios-Navarro G, Fardoun HM, et al. Virtual rehabilitation for multiple sclerosis using a kinect-based system: randomized controlled trial. JMIR Serious Games 2014 Nov 12;2(2):e12 [FREE Full text] [doi: 10.2196/games.2933] [Medline: 25654242]
25. Trost Z, Zielke M, Guck A, Nowlin L, Zakhidov D, France CR, et al. The promise and challenge of virtual gaming technologies for chronic pain: the case of graded exposure for low back pain. Pain Manag 2015;5(3):197-206. [doi: 10.2217/pmt.15.6] [Medline: 25971643]
26. Kerr KM. Movement science: foundations for physical therapy in rehabilitation. Physiotherapy 1990 Feb;76(2):111. [doi: 10.1016/s0031-9406(10)62529-1]
27. Adams DL. Develop better motor skill progressions with gentile's taxonomy of tasks. J Phy Educ Recr Dance 1999 Oct;70(8):35-38. [doi: 10.1080/07303084.1999.10605704]

28. Thomas JS, France CR, Applegate ME, Leitkam ST, Walkowski S. Feasibility and safety of a virtual reality dodgeball intervention for chronic low back pain: a randomized clinical trial. J Pain 2016 Dec;17(12):1302-1317 [FREE Full text] [doi: 10.1016/j.jpain.2016.08.011] [Medline: 27616607]
29. France CR, Thomas JS. Virtual immersive gaming to optimize recovery (VIGOR) in low back pain: a phase II randomized controlled trial. Contemp Clin Trials 2018 Jun;69:83-91 [FREE Full text] [doi: 10.1016/j.cct.2018.05.001] [Medline: 29730393]
30. ISO 9241-11:1998 Ergonomic Requirements for Office Work With Visual Display Terminals (Vdts) — Part 11: Guidance on Usability. ISO - International Organization for Standardization. 1998. URL: https://www.iso.org/standard/16883.html [accessed 2020-07-02]
31. Kelley ML, Heffer RW, Gresham FM, Elliott SN. Development of a modified treatment evaluation inventory. J Psychopathol Behav Assess 1989 Sep;11(3):235-247. [doi: 10.1007/bf00960495]
32. Peshkin M. KineAssist: a Robotic Overground Gait and Balance Training Device. In: Proceedings of the 9th International Conference on Rehabilitation Robotics. 2005 Presented at: ICORR'05; June 28-July 1, 2005; Chicago, IL, USA. [doi: 10.1109/icorr.2005.1501094]
33. Patton J, Brown DA, Peshkin M, Santos-Munné JJ, Makhlin A, Lewis E, et al. KineAssist: design and development of a robotic overground gait and balance therapy device. Top Stroke Rehabil 2008;15(2):131-139. [doi: 10.1310/tsr1502-131] [Medline: 18430678]
34. Wang J, Hurt CP, Capo-Lugo CE, Brown DA. Characteristics of horizontal force generation for individuals post-stroke walking against progressive resistive forces. Clin Biomech (Bristol, Avon) 2015 Jan;30(1):40-45 [FREE Full text] [doi: 10.1016/j.clinbiomech.2014.11.006] [Medline: 25481856]
35. Hurt CP, Wang J, Capo-Lugo CE, Brown DA. Effect of progressive horizontal resistive force on the comfortable walking speed of individuals post-stroke. J Neuroeng Rehabil 2015 Mar 10;12:12 [FREE Full text] [doi: 10.1186/s12984-015-0007-7] [Medline: 25884953]
36. Rumble DD, Hurt CP, Brown DA. Step-by-step variability of swing phase trajectory area during steady state walking at a range of speeds. PLoS One 2018;13(1):e0191247 [FREE Full text] [doi: 10.1371/journal.pone.0191247] [Medline: 29370202]
37. Graham SA, Hurt CP, Brown DA. Minimizing postural demands of walking while still emphasizing locomotor force generation for nonimpaired individuals. IEEE Trans Neural Syst Rehabil Eng 2018 May;26(5):1003-1010. [doi: 10.1109/TNSRE.2018.2810701] [Medline: 29752235]
38. Dionisio VC, Brown DA. Collaborative robotic biomechanical interactions and gait adjustments in young, non-impaired individuals. J Neuroeng Rehabil 2016 Jun 16;13(1):57 [FREE Full text] [doi: 10.1186/s12984-016-0166-1] [Medline: 27306027]
39. Kugler K, Wijn J, Geilen M, de Jong J, Vlaeyen J. The photograph series of daily activities (PHODA). In: Pain-Related Fear:Exposure-Based Treatment of Chronic Pain: Exposure-Based Treatment. North York, ON: Wolters Kluwer; 1999.
40. Trost Z, France CR, Thomas JS. Exposure to movement in chronic back pain: evidence of successful generalization across a reaching task. Pain 2008 Jul;137(1):26-33. [doi: 10.1016/j.pain.2007.08.005] [Medline: 17869423]
41. Roelofs J, Goubert L, Peters ML, Vlaeyen JW, Crombez G. The Tampa Scale for Kinesiophobia: further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. Eur J Pain 2004 Oct;8(5):495-502. [doi: 10.1016/j.ejpain.2003.11.016] [Medline: 15324781]
42. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, IL: Human Kinetics; 1998.
43. Demoulin C, Verbunt JA, Winkens B, Knottnerus JA, Smeets RJ. Usefulness of perceived level of exertion in patients with chronic low back pain attending a physical training programme. Disabil Rehabil 2010;32(3):216-222. [doi: 10.3109/09638280903071842] [Medline: 20001827]
44. Amano S, Ludin AF, Clift R, Nakazawa M, Law TD, Rush LJ, et al. Effectiveness of blood flow restricted exercise compared with standard exercise in patients with recurrent low back pain: study protocol for a randomized controlled trial. Trials 2016 Mar 12;17:81 [FREE Full text] [doi: 10.1186/s13063-016-1214-7] [Medline: 26867541]
45. Penn T, Browning W, France C, Hardee G, Zielke M, Trost Z. Attitudes toward a virtual reality physical activity intervention among veterans with chronic low back pain. J Pain 2017 Apr;18(4):S1 [FREE Full text] [doi: 10.1016/j.jpain.2017.02.004]
46. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
47. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
48. Melzack R. The short-form McGill pain questionnaire. Pain 1987 Aug;30(2):191-197. [doi: 10.1016/0304-3959(87)91074-8] [Medline: 3670870]
49. Goubert L, Crombez G, Van Damme S, Vlaeyen JW, Bijttebier P, Roelofs J. Confirmatory factor analysis of the Tampa Scale for Kinesiophobia: invariant two-factor model across low back pain patients and fibromyalgia patients. Clin J Pain 2004;20(2):103-110. [doi: 10.1097/00002508-200403000-00007] [Medline: 14770050]
50. Wright KD, Asmundson GJ, McCreary DR. Factorial validity of the short-form McGill pain questionnaire (SF-MPQ). Eur J Pain 2001;5(3):279-284. [doi: 10.1053/eujp.2001.0243] [Medline: 11558983]
51. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: visual analog scale for pain (VAS Pain), numeric rating scale for pain (NRS Pain), McGill pain questionnaire (MPQ), short-form McGill pain questionnaire (SF-MPQ),

chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]

52. Olaogun MO, Adedoyin RA, Ikem IC, Anifaloba OR. Reliability of rating low back pain with a visual analogue scale and a semantic differential scale. Physiother Theor Pr 2009 Jul 10;20(2):135-142. [doi: 10.1080/09593980490453048]
53. Yelvar GD, Çırak Y, Dalkılınç M, Demir Y, Guner Z, Boydak A. Is physiotherapy integrated virtual walking effective on pain, function, and kinesiophobia in patients with non-specific low-back pain? Randomised controlled trial. Eur Spine J 2017 Feb;26(2):538-545. [doi: 10.1007/s00586-016-4892-7] [Medline: 27981455]
54. Fowler CA, Ballistrea LM, Mazzone KE, Martin AM, Kaplan H, Kip KE, et al. Virtual reality as a therapy adjunct for fear of movement in veterans with chronic pain: single-arm feasibility study. JMIR Form Res 2019 Oct 30;3(4):e11266 [FREE Full text] [doi: 10.2196/11266] [Medline: 31670696]
55. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]

Abbreviations

ANOVA: analysis of variance
cLBP: chronic low back pain
FAM: fear-avoidance model
GEXP: graded exposure
HMD: head-mounted display
KA-MX: KineAssist-MX
MPQ-sf: McGill pain questionnaire-short form
PRI: pain rating index
RPE: rating of perceived exertion
SUS: system usability scale
TEI-sf: treatment evaluation inventory-short-form
TSK: Tampa scale for kinesiophobia
VAS: visual analog scale
VR: virtual reality

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by S Graham, A Garcia, D Lipardo, A Rehan Youssef; comments to author 30.03.20; revised version received 05.05.20; accepted 03.06.20; published 10.08.20.

Please cite as:
Hennessy RW, Rumble D, Christian M, Brown DA, Trost Z
A Graded Exposure, Locomotion-Enabled Virtual Reality App During Walking and Reaching for Individuals With Chronic Low Back Pain: Cohort Gaming Design
JMIR Serious Games 2020;8(3):e17799
URL: http://games.jmir.org/2020/3/e17799/
doi:10.2196/17799
PMID:32773381

Original Paper

Serious Game on a Smartphone for Adolescents Undergoing Hemodialysis: Development and Evaluation

Cristina Célia De Almeida Pereira Santana[1*], MS; Ana Tereza Vaz De Souza Freitas[2*], PhD; Gilson Oliveira Barreto[3*], PhD; Igor Sousa De Avelar[3*], BA; Renata Mazaro-Costa[4*], PhD; Gina Nolêto Bueno[5*], PhD; Diuly Caroline Ribeiro[6*], BA; Gabriela Damasceno Silva[6*], BA; Alessandra Vitorino Naghettini[6*], PhD

[1]Hospital das Clínicas, Postgraduate Program Teaching in Health, Goias Federal University, Goiânia, Brazil
[2]Faculty of Nutrition, Goias Federal University, Goiânia, Brazil
[3]Laboratory of Information Technology, Goias Federal University, Goiânia, Brazil
[4]Institute of Biological Sciences, Goias Federal University, Goiânia, Brazil
[5]Departament of Psychology, Pontifical Catholic University of Goiás, Goiânia, Brazil
[6]Faculty of Medicine, Postgraduate Program Teaching in Health, Goias Federal University, Goiânia, Brazil
[*]all authors contributed equally

Corresponding Author:
Cristina Célia De Almeida Pereira Santana, MS
Hospital das Clínicas
Postgraduate Program Teaching in Health
Goias Federal University
Primeira Avenida, S/nº, Setor Universitário
Goiânia, 74605-020
Brazil
Phone: 55 62991507049
Email: ccaps44@gmail.com

Abstract

Background: Adolescents with chronic kidney disease have a hard time adhering to hemodialysis as a therapy, indicating a need to establish new alternatives for motivation and adherence to treatment.

Objective: The objective of this study was to develop and evaluate a serious game to stimulate and motivate adolescents undergoing hemodialysis.

Methods: We describe the technological production followed by a qualitative analysis. We invited 8 adolescents undergoing hemodialysis in the city Goiânia, located in the midwest of Brazil, to participate. The final convenience sample included 7 (87.5% of the target population) adolescents. The process was conducted in 3 phases: creation of a serious game, evaluation of its use, and observation of its motivating effect on behavioral modification with a focus on acquiring the necessary competence for self-care.

Results: An app (Bim) in the modality of a serious game was developed to be used during hemodialysis; the player was encouraged to take care of a character with daily actions during his or her treatment. The game was made available to adolescents aged 10-14 years. Mobile devices were offered during the hemodialysis treatment for a period of 30-40 minutes, 3 times a week for 60 days. The usage definitions of the game were freely chosen by the participants. The qualitative evaluation of the use of the Bim app showed that it encompasses scenarios and activities that enable the exercise of daily actions for the treatment of patients. The behavioral evaluation showed that the Bim app worked as a motivating stimulus for behavioral adherence to hemodialysis requirements.

Conclusions: The easy-to-access app interface showed good operability for its users. The description of the character and proposed activities contributed to motivation and ability to cope with hemodialysis care.

(JMIR Serious Games 2020;8(3):e17979) doi:10.2196/17979

KEYWORDS

adolescent; hemodialysis; serious game; operability

Introduction

Chronic kidney disease (CKD) is a major public health problem worldwide. It is estimated that the prevalence of renal replacement therapy ranges from 18 per million to 100 per million in youth under 18 years of age [1]. In Brazil, approximately 700 young people under 19 years of age undergo hemodialysis, according to the 2017 Census by the Sociedade Brasileira de Nefrologia [2].

Hemodialysis patients are subject to extensive changes in style and quality of life due to arduous clinical and therapeutic control of the disease, in addition to recurrent hospitalizations due to complications [3,4].

The impact is greater among adolescents and negatively affects their growth and development, in addition to contributing to family disorganization, social exclusion, and the prospect of death, which simultaneously affect their families and caregivers [5,6].

With the difficulty in adhering to hemodialysis comes the need to implement new strategies to facilitate care planning and adaptation to hemodialysis treatment, in order to optimize results in self-care promotion, one of the main axes of health care for this population [5-7].

The use of games has been proposed as an effective method for behavioral changes, influencing the results [8-10] and considered as a potentially effective health intervention for treatment adherence since it motivates self-care in adolescents by the means of greater access to information on the disease [11-13].

eHealth has been reaching users of all social classes and age groups, enabling the development of tools such as serious games and digital media focused on health educational processes [8,11,13]. A study published in 2014 presented a serious game developed to stimulate care in children with hemophilia [14]; however, there have been no games in this modality aimed at adolescents undergoing hemodialysis yet [15].

Therefore, the objective of this study was to develop a serious game that would encourage adolescents undergoing hemodialysis to perform self-care and adhere to the treatment.

Methods

Outline

This study involved the technological production of the serious game followed by a qualitative analysis.

An app in the modality of a serious game was developed with a simulator and was complemented by a series of mini-game actions to be used by adolescents under hemodialysis; the use, motivational capacity, and game interaction of the app were evaluated.

The process was conducted in 3 phases: creation of a serious game, its evaluation while being used, and observation of the motivating effect for behavioral modification, with a focus on acquiring the necessary competence for self-care regarding the adherence to medical nutritional instructions appropriate to the condition of a dialysis patient.

Sample

The population consisted of patients aged 10-14 years undergoing hemodialysis in the city of Goiânia, in the midwest of Brazil. The 8 patients undergoing hemodialysis in the city were personally invited by the researcher. They were all instructed about the study proposal and how it would be carried out, and an adult responsible for each participant signed a term of consent. One of the adolescents did not participate as he was transferred to another state for a kidney transplant.

The final study sample was defined by convenience and included 7 adolescents (87.5% of the target population). Participants did not need internet access to play the serious game since the app was already installed on the tablets distributed to each of them.

The exclusion criteria included refusal to participate in the study and the presence of any cognitive deficit that could possibly hinder the proper use of the app.

This study was given a favorable opinion by the Ethics Committee on Human and Animal Medical Research of the Clinical Hospital, Federal University of Goiás, under the consubstantiated opinion number 1.455.896 (CAAE: 53877316.9.0000.5078).

Study Phases

Phase 1: The Creation Process of a Serious Game

The app point of origin was the idea of using the Tamagotchi device that has been used by children since 1990 and bringing it to the universe of CKD. Other highlights are the applied 3-dimensional technology and visual identity of the character referring to a kidney that includes being red and having playful and easily identifiable features.

The conceptual development of the Bim app was supported by applied behavior analysis, which considers that the way individuals describe the events with which they interact changes their emotional response, which will in turn affect the behavior of either approaching (joining) those events or walking away or even fighting against them (not joining). Therefore, it is science that describes the role of individuals' behavior [16,17].

The experiential gaming model was used to reflect on the educational design. This model proposes a sequence of elements that allow players to think over the knowledge acquired and evaluate their own performance while handling the app, which characterizes the activities as cognitive-behavioral [18].

Seeking to promote a sense of commitment in adolescents must be considered, using a contextualized and positive perspective and the capability of addressing diverse situations [19].

In the game, the graphics and icons are personalized, and immersion is favored by the interaction with scenarios, everyday situations, and caring for the character. Learning is achieved in the understanding of health needs by the means of a scoring system and a simulated response in the virtual environment.

The serious game was designed for tablets and smartphones using the Android operating system, and the Adobe Photoshop CS6, Adobe Illustrator CS6, Maya 3D, and Unity 3D programs were used. Product development was carried out at the

Laboratory of Technologies and Media for Education at the Federal University of Goiás (LabTIME/UFG).

Phase 2: Evaluation of the Use of the Serious Game

During hemodialysis sessions, adolescents and their companions were informed about the research and invited to participate. After providing written, informed consent, tablets with the app installed were made available.

The Bim app was installed on a 7-inch Samsung Galaxy Tab Tablet model E T113, with 8GB of memory, Android 4.4 system, and Quad Core processor with 1.3 GHz. This configuration, according to developers, guarantees better operability.

Mobile devices were lent to the adolescents during hemodialysis treatment for a period of 30-40 minutes, 3 times a week for 60 days. The game usage definitions were freely chosen by the participants. Two resources were used to evaluate the use of the game: semistructured interviews and onsite observation of use.

The semistructured interviews were conducted with the users after the app availability period. We used 7 guiding questions to identify the motivation to use the app and the perception of care actions for the character. It is noteworthy that caregivers were also approached, to investigate their perceptions about the app and identify possible interfaces of use with the performance of self-care by the adolescents in the domestic or hospital environments. The interviews lasted around 20 minutes and were carried out according to the participants' availability in the hemodialysis environment itself. They were recorded in mp4 audio and transcribed for analysis afterwards.

A weekly onsite evaluation was conducted to observe the use of the app, totaling 8 observations while the tablets were available. We selected 3 guiding criteria for observation to identify the handling of resources, interest or motivation while performing the activities, and the occurrence of favorable or opposite situations in the self-care learning process provided by the game usage.

The impressions identified in this step were recorded in a field diary.

The data obtained during the semistructured interviews and in the observation sessions were analyzed using content analysis, in the form of thematic analysis. The steps used for the data treatment were guided by Bardin [20], which allows systematizing the nuclei of meaning of the participants' responses. The participants in this study were organized by letters and numbers, according to the order in which they were interviewed (A-adolescent).

Phase 3: Behavioral Evaluation

The behavioral evaluation phase of this study was conducted using applied behavior analysis methodology, which encompasses behavioral science and seeks to observe and analyze the effects of the use of the game on the behavioral pattern of the participants, focusing on the increase in adherence to the treatment demanded by the disease. In this sense, the verbal and nonverbal responses given during the use of the game by the participants in this study were taken into consideration; these responses were related to their capacity to discriminate the effects of adequate or inadequate care given to the character and its correlation to their own behavioral pattern in the face of the disease and hemodialysis treatment they underwent.

Results

Serious Game Creation Process

The creation and conception process (Figure 1) required meetings with the multidisciplinary team, which included a graphic designer, game designer, nutritionist, pediatrician, nurse, biologist, educator, and psychologist. During these meetings, both the interface and all the necessary aspects of the game such as the Booleans, parameters, and environments were discussed.

This version of the serious game offers a virtual environment that includes a home and space for dialysis treatment (Figure 2). The graphic orientation intends to stimulate task management, the ability to pay attention, and the ability to care for the character. A box with a tag (presentation letter) is used to activate it, and the tag reads (Figure 2C): "My name is Bim. Do you wanna be my friend? I like to play, eat and receive a lot of affection. I just can't take care of myself. I have a problem called chronic kidney disease. I can drink a little water, and I need to take medicine every day. You know what I'm talking about, right? That must be why they brought me to you."

Figure 1. Process of creating the character, which resembles a kidney. The various expressions are meant to corroborate with the player's perception of the character's emotions while performing an action in the game.

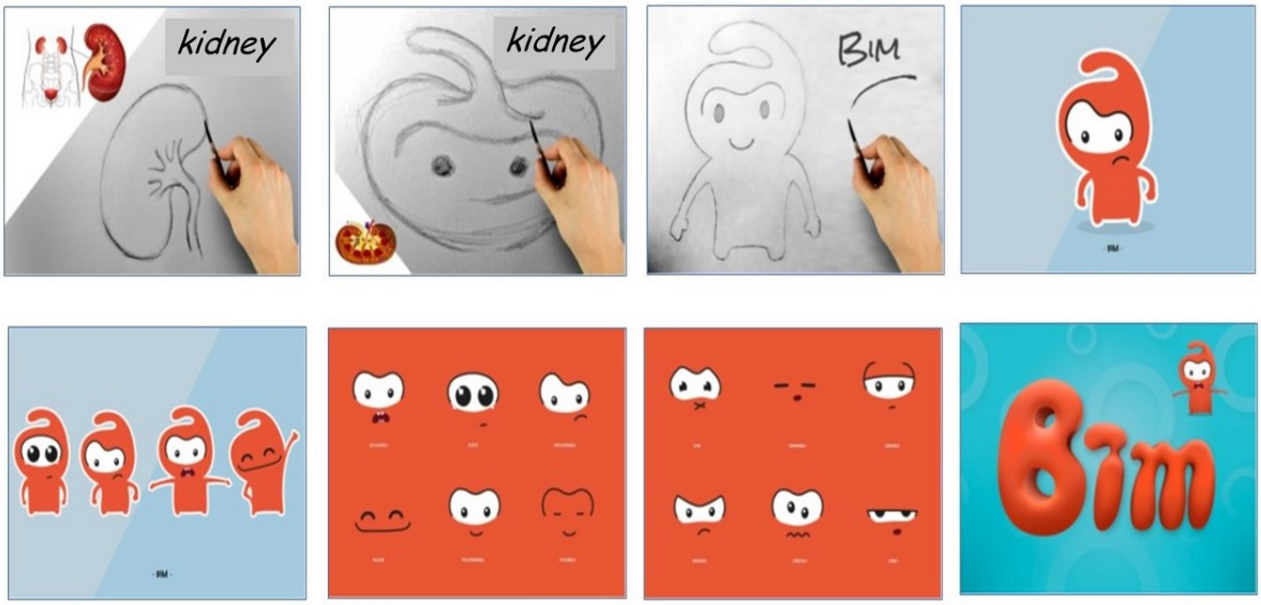
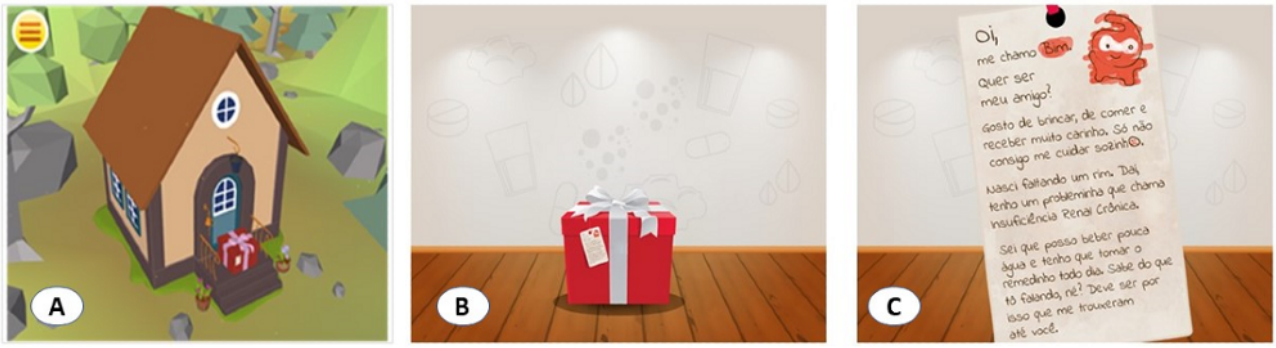


Figure 2. The initial app design. A: home scenario; B: gift box; C: presentation letter.



As soon as the game starts, the adolescent is encouraged to take care of the character with daily actions that are experienced in his daily life and health condition such as: food and water intake, hygiene, use of medication, and dialysis session in several scenarios alluding to the domestic and hospital environments (Figure 3).

The home screen contains self-explanatory icons. It also gives access to a tutorial that enables the user to supplement information about the use and other features of the app. The actions are guided by icons next to the screen that indicate, with colors, the need for care or its excess (Figures 4 and 5) (Multimedia Appendix 1).

The game parameters simulate the 24 hours in a day and include the following in its normal cycle: 6 meals, 3 opportunities for water intake, 4 opportunities to use medication, 1 hemodialysis session, 1 need for sleep or rest, and 2 opportunities for body hygiene. The parameters are signaled by the following colors: green (normal), yellow (alert), purple (lacking care), and black (excessive care). However, the player is free to perform a sequence of care.

Each activity performed by the player is mediated by a score, and each action that is considered positive or assertive is given a higher score. The positive balance of points allows the player to accumulate crystal bonuses that can be used to access mini games that stimulate motor coordination and memory. The additional gameplay provides an opportunity for more learning associated with leisure (Figure 6). Musical features and the character's interactive facial expressions are present in all the bonus games.

Situations in which the avatar, in the case of Bim, was not properly cared for could lead to its hospitalization. Once hospitalized, the character remains without access to any action for 30 seconds, as shown in Figure 7. It was expected that, during this period, the player would think about the action taken with the character, which he should have taken care of, considering the prescriptions for its health condition. Afterwards, the parameters were normalized, the game was restarted, and the accumulated bonuses were lost.

Figure 3. Scenarios in the serious game. A: living room; B: kitchen; C: bedroom; D: bathroom; E: hemodialysis room; F: hospital inpatient bed. At the sides, the scenarios display commands (icons and colors) that guide or indicate the proper care to guarantee a higher score.

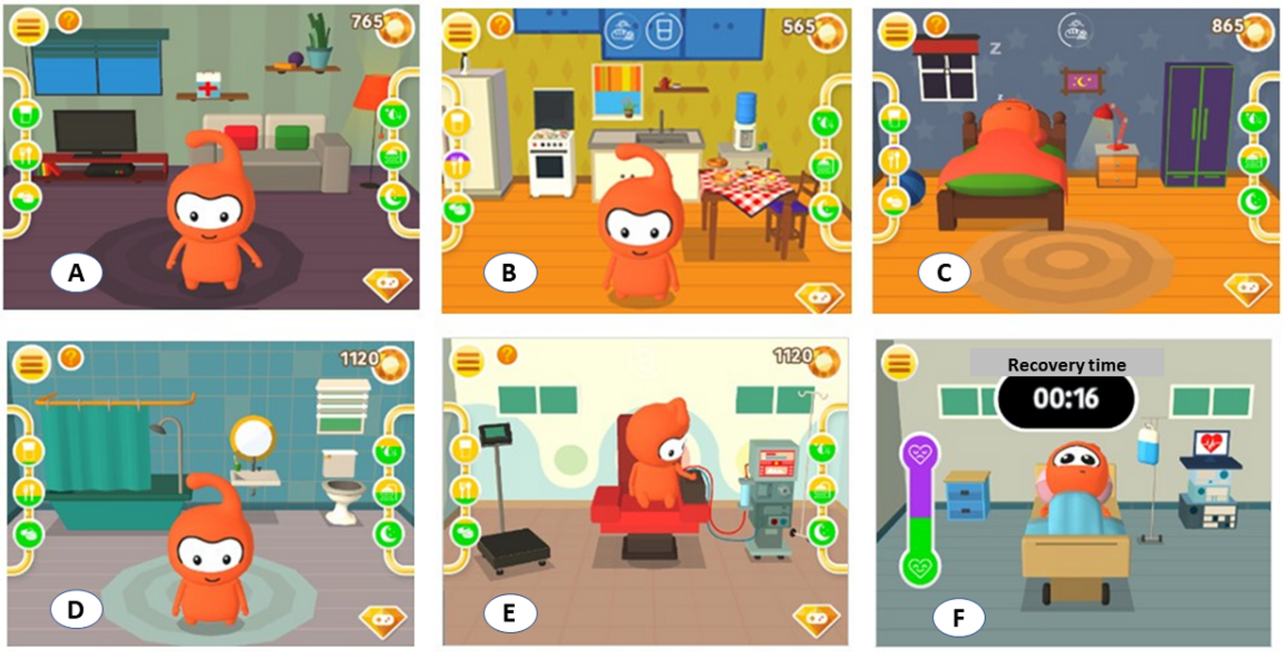
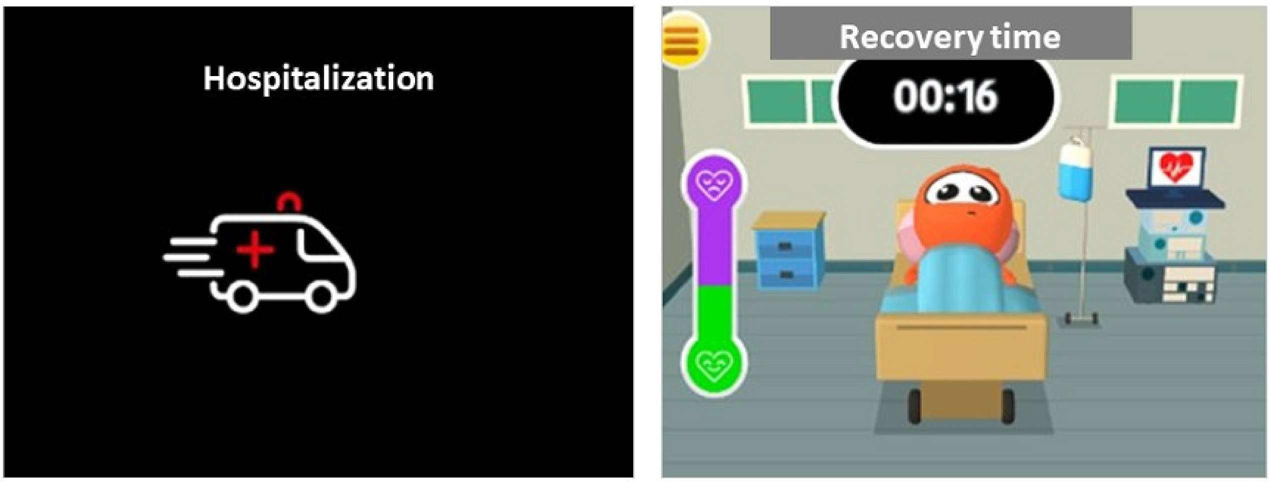


Figure 4. Main app interface and game parameters. Note the icons for accessing the menu, mini games, and tutorial; for viewing points; and suggesting the need for care: thirst (glass with water), hunger (fork and knife), medicine (pills), hemodialysis (blood drop and arrows), hygiene (shower), and fatigue (moon and stars). The icon colors green (normal parameter), yellow (warning sign), purple (lacking care), and black (excessive care) allow combinations where hospitalization conditions are associated with lack or excess of care and the release of points to access bonus mini games.

Figure 5. Instructions for the parameters and conditions for hospitalization and the access to bonus games.

Figure 6. Bonus mini games. A: home screen; B: BIMCEU, a mini-game that stimulates motor coordination by ascending platforms and avoiding obstacles; C: BIMSOM, which requires the repetition of an increasingly complex sequence of sounds and colors until the final level is reached; D: BIMPOTE, which encourages memory while searching for the character.

Figure 7. Hospitalization alert symbolized by an ambulance and the infirmary setting. The thermometer icon on the left side of the character's bed uses color to indicate the recovery and discharge time.



Evaluation of the Use of the Serious Game

For this evaluation, 7 adolescents were observed, aged 10-14 years: 2 (28.6%) aged 10 years, 1 (14.3%) aged 12 years, 3 (42.8%) aged 13 years, and 1 (14.3%) aged 14 years. Of the 7 participants, 3 (42.8%) were female, and 4 (57.2%) were male. Time with CKD ranged from 6 months to 10 years; 4 (4/7, 57.2%) had been diagnosed with CKD less than 2 years ago. In the entire sample, hemodialysis treatment had started less than 3 years earlier; 5 (5/7, 71.4%) adolescents had been undergoing hemodialysis for up to 1 year. The average time of the dialysis session was 3.5-4.0 hours, 3 times a week.

In the qualitative analysis of the responses obtained from the interviews, the following categories were identified: use, motivation regarding the use of the serious game, perception of the clinical condition of the character, and care actions performed in the serious game.

The participants showed interest in the serious game, which could be verified by the interaction and motivation that led to the fulfillment of the activities proposed.

The game handling was considered simple by the participants. There was a good understanding of the main panel, easy access to scenarios, quick association of colors to the lacking of or excessive care for the character, and the perception that assertive care generated crystal bonuses and their exchange gave access to mini games (BIMCEU, BIMSOM, and BIMPOTE).

The most accessed game was BIMPOTE, and the least accessed was BIMSOM, which was quite burdensome when it came to memorization and achieving more complex levels. Table 1 shows the evaluated items and the adolescents' impressions on the use and motivation regarding the game.

Table 1. Evaluation of the use and motivation to handle the Bim app.

Parameters and requirements	Response (n=7), n (%)	
	Yes	No
Use of the game		
Difficulty while playing	0 (0)	7 (100)
Good handling of the main panel	5 (71.4)	2 (28.6)
Understanding the meaning of icons	7 (100)	0 (0)
Understanding the meaning of colors	6 (85.7)	1 (14.3)
I learned from the game	7 (100)	0 (0)
Most accessed mini game: BIMPOTE	4 (57.1)	N/A[a]
Least accessed mini game: BIMSOM	5 (71.4)	N/A
High number of BIM hospitalizations	2 (28.6)	5 (71.4)
Tablet crashed while playing	1 (14.3)	6 (85.7)
Motivation		
I liked the character Bim	7 (100)	0 (0)
I liked to take care of Bim	7 (100)	0 (0)
I liked the colors and music in the game	7 (100)	0 (0)
I would like to continue taking care of Bim	6 (85.7)	1 (14.3)

[a]N/A: not applicable.

Analysis of the content of the responses also signaled that the virtual reality proposed by the game enabled perception of the character's clinical condition, which was similar to the adolescents' realities:

> *It is a good game showing our reality, of those who have kidney failure. I thought it was cool for him to undergo hemodialysis.* [A1]

> *It's [the character's routine] normal, just like mine.* [A5]

The game, as a playful activity, revealed potential for promoting the learning of necessary actions for self-care, providing the practice of the guidelines given by health professionals and family members:

> *He cannot drink a lot of water or he [the character] might be swollen for hemodialysis.* [A6]

> *I'm already used to it ... because I know how hemodialysis is ... how my diet is.* [A2]

The hospitalization context appeared to be commonly associated with nonadherence to food care:

> *A little difficult! Because he is going to the hospital! He wants to eat something, but he can't!* [A4]

Despite this, it allowed the association of unfulfilled care with the emerged complication:

> *I learned that if don't take the medicine, we end up at the hospital.* [A1]

> *I learned that we have to keep a proper diet; otherwise, we'll get sick and be hospitalized!* [A7]

On the other hand, it was observed that the character's hospitalization was perceived with curiosity by the youngest participant, who repeated the nonassertive care to prove the effects of damage on the character. When she was questioned why the character was always being hospitalized, A3 replied: "I gave him a lot of water!" This adolescent was unable to adhere to water restrictions in her daily life, as reported.

Behavioral Evaluation

The verbal reports obtained from the evaluation of the use showed that the serious game was a motivating stimulus (Sr+: positive reinforcing stimulus) not only for the patient-participants (A1, A2, A5, A7), as highlighted, but also for their caregivers and even the assisting professionals who were able to interact with a technology that taught the patients instructions they wanted them to learn and practice. The patient-participants noticed that the character's health condition was similar to theirs; they were also able to observe the consequence of the emitted behaviors, a *sine qua non* condition for behavioral modification, which implies a wide learning process. They discriminated (Sd: discriminative stimulus) that incorrectly medicating the character (R: response to the condition of a patient with CKD undergoing hemodialysis) led to hospitalization, as well as releasing (R) a larger amount of water meant that he arrived to hemodialysis with very altered wet weight (C: aversive consequence generated by the opposite behavior to what is instructed to patients under treatment for this disease). Sensitization of the caregivers while monitoring the adolescents during the operationalization of the serious game was also observed: This made the adolescent watch the effect on the character and understand what not following the treatment instructions produced in his own organism, as highlighted by the caregiver of A4:

> *That's what she goes through! She's there playing something she lives in day to day, right? She realizes what is going on in the game, nothing better, right?* [caregiver of A4]

A relevant description was also presented by the caregiver of A1:

> *Nowadays, for their age group, it's all about game, it's all about the internet. (...) what is really going to affect them is right there.* [caregiver of A1]

They learn by the means of tools, that is to say, "teaching machines," or by the means of the serious game, "game or internet."

Discussion

Principal Findings

In this study, we developed a serious game to be used by adolescents during hemodialysis sessions. This app has a straightforward interface and guidelines for game usage. Its use during the session provided the adolescent with an experience that allowed both the routine and duration of the treatment to be reduced. The activities proposed by the game also contributed to the biopsychosocial aspects and their orientation towards self-care.

Recent studies show the potential of serious games in improving young people's health outcomes. However, few refer to its use by adolescents undergoing hemodialysis [21]. In this scenario, the individual's interaction with virtual and multimedia resources, such as interactive games, works as a reinforcing contingency to favor learning [8,10,22-24].

The expansion and use of technology are related to overcoming mobility barriers [25]. The young population has shown good receptivity to educational electronic media, as they already use the internet and other technological resources daily [26-28].

A meta-analysis on the promotion of a healthy lifestyle by serious games has shown a positive effect; the most significant benefits were the understanding of clinical results and the maintenance of healthy long-term behaviors [29].

The insertion of serious games in both hospital and domestic contexts can lead to immersion in the experienced issue, providing an opportunity to discuss the conditions for health promotion, in addition to promoting reflection on unhealthy habits or inefficient care [14,26,28,30,31].

As far as we know, this is the first study that developed an avatar that proposed the learning of self-care focused on adolescents undergoing hemodialysis. The behavioral evaluation guided the presence of the reinforcing stimulus, which motivated the adolescents, their companions, and health professionals, and they could observe the consequence of the performed behaviors.

Games can improve the availability of reinforcing contingencies and establish motivating contexts for the modeling of necessary behavioral topographies, such as those required to accomplish the treatment of patients with CKD [21,32].

The use of serious games with playful guidance can appropriately guide the care of adolescents on hemodialysis. That said, a long-term evaluation with a larger number of patients may demonstrate interference (favoring adherence to treatment) in clinical aspects. The availability of games for caregivers can also reinforce the desired behavior change.

However, it is worth noting that even though it is an unprecedented device — a serious game for adolescents with CKD — it is still the first version and therefore needs to be optimized. In this regard, the present study opened new research horizons with the Bim app, in terms of a game, improving its navigation, improving the warnings and reminders to the user, increasing the interaction interface between the avatar and player, seeking to broaden humanization, and expanding the mini games.

When it comes to the users and caregivers, both training and engaging activities should be developed in the app; consequently, the app should be turned into part of a digital platform where both the patient and caregivers can be offered important information on the disease and how to better manage it. Expanding its perspective, the Bim app could be programed for other avatars with other chronic diseases such as obesity and cystic fibrosis.

Conclusion

The results measured in this study emphasize the lived experience for the development and evaluation of an interactive app, in the serious game modality, to be used by patients undergoing hemodialysis. The interface proved to be a tool that contributes to the motivation and reinforcement of the care that adolescents need during treatment. Therefore, these results are expected to expand current knowledge about the effectiveness of mobile health interventions in renal replacement therapy so that this tool can be used by this population.

Acknowledgments

We thank the Information Technology and Educational Media Laboratory of the Institute of Biological Sciences of the Federal University of Goiás and FLIP Educational Technologies. RMC was a fellow of PET-SESu-MEC. PROEXT Announcements 2015 and 2016. CS was supported by the Goiás Research Support Foundation (FAPEG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bim app video.
[MP4 File (MP4 Video), 26993 KB - games_v8i3e17979_app1.mp4]

References

1. Harambat J, van Stralen KJ, Kim J, Tizard E. Epidemiology of chronic kidney disease in children. Pediatr Nephrol 2012 Mar 27;27(3):363-373 [FREE Full text] [doi: 10.1007/s00467-011-1939-1] [Medline: 21713524]
2. Thomé FS, Sesso RC, Lopes AA, Lugon J, Martins C. Brazilian chronic dialysis survey 2017. J Bras Nefrol 2019;41(2):208-214 [FREE Full text] [doi: 10.1590/2175-8239-JBN-2018-0178] [Medline: 30968930]
3. Fadrowski J, Cole SR, Hwang W, Fiorenza J, Weiss RA, Gerson A, et al. Changes in physical and psychosocial functioning among adolescents with chronic kidney disease. Pediatr Nephrol 2006 Mar;21(3):394-399. [doi: 10.1007/s00467-005-2122-3] [Medline: 16382317]
4. McKenna A, Keating L, Vigneux A, Stevens S, Williams A, Geary D. Quality of life in children with chronic kidney disease-patient and caregiver assessments. Nephrol Dial Transplant 2006 Jul;21(7):1899-1905. [doi: 10.1093/ndt/gfl091] [Medline: 16611686]
5. Ministério da Saúde. Diretrizes clínicas para o cuidado ao paciente com doença renal crônica ? DRC no Sistema Único de Saúde. Secretaria de Atenção à Saúde, Departamento de Atenção Especializada e Temática. Brasília: Ministério da Saúde; 2014. URL: http://bvsms.saude.gov.br/bvs/publicacoes/diretrizes_clinicas_cuidado_paciente_renal.pdf [accessed 2020-09-04]
6. Campos C, Mantovani MF, Nascimento M, Cassi CC. [Social representations of illness among people with chronic kidney disease]. Rev Gaucha Enferm 2015 Jun;36(2):106-112 [FREE Full text] [doi: 10.1590/1983-1447.2015.02.48183] [Medline: 26334416]
7. The global diffusion of eHealth: making universal health coverage achievable report of the third global survey on eHealth. World Health Organization. 2016 Dec. URL: https://www.who.int/goe/publications/global_diffusion/en/ [accessed 2020-09-04]
8. Clochesy J, Buchner M, Hickman R, Pinto M, Znamenak K. Creating a Serious Game for Health. J Health Hum Serv Adm 2015;38(2):162-173. [Medline: 26442360]
9. Ingadottir B, Blondal K, Thue D, Zoega S, Thylen I, Jaarsma T. Development, Usability, and Efficacy of a Serious Game to Help Patients Learn About Pain Management After Surgery: An Evaluation Study. JMIR Serious Games 2017 May 10;5(2):e10 [FREE Full text] [doi: 10.2196/games.6894] [Medline: 28490419]
10. Laamarti F, Eid M, El Saddik A. An Overview of Serious Games. International Journal of Computer Games Technology 2014 Jan;2014:1-15 [FREE Full text] [doi: 10.1155/2014/358152]
11. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin L, Green C, et al. Games for Health for Children-Current Status and Needed Research. Games Health J 2016 Feb 5;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
12. Staiano AE, Flynn R. Therapeutic Uses of Active Videogames: A Systematic Review. Games Health J 2014 Dec;3(6):351-365 [FREE Full text] [doi: 10.1089/g4h.2013.0100] [Medline: 26192642]
13. Payne H, Lister C, West J, Bernhardt J. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
14. Matsunaga R, Moraes R, Borges M, Matta M, Ozelo M. Development of a serious game for children with hemophilia. J Health Inform 2014;6:114-119 [FREE Full text] [doi: 10.1109/segah.2014.7067082]
15. Santana C, Naghettini A, Freitas A, Barreto G, Mazaro-Costa R. Aplicativos como estratégia de ensino na doença renal crônica infantil: uma revisão da literatura. J Health Inform 2016;8:287-297 [FREE Full text]
16. Starks K. Cognitive behavioral game design: a unified model for designing serious games. Front Psychol 2014;5:28 [FREE Full text] [doi: 10.3389/fpsyg.2014.00028] [Medline: 24550858]
17. Skinner BF. Science and human behavior. New York, NY: Free Press; 2014.

18. Kiili K. Content creation challenges and flow experience in educational games: The IT-Emperor case. The Internet and Higher Education 2005 Jul;8(3):183-198 [FREE Full text] [doi: 10.1016/j.iheduc.2005.06.001]
19. Zappe J, Dell'Aglio DD. Adolescência em diferentes contextos de desenvolvimento: risco e proteção em uma perspectiva longitudinal. Psico 2016 May 12;47(2):99-110 [FREE Full text] [doi: 10.15448/1980-8623.2016.2.21494]
20. Bardin L. Análise de Conteúdo. São Paulo: Edições 70; 2016.
21. Holtz B, Murray K, Park T. Serious Games for Children with Chronic Diseases: A Systematic Review. Games Health J 2018 Oct;7(5):291-301 [FREE Full text] [doi: 10.1089/g4h.2018.0024] [Medline: 29957082]
22. Boendermaker W, Veltkamp R, Peeters M. Training Behavioral Control in Adolescents Using a Serious Game. Games Health J 2017 Dec 6;6(6):351-357 [FREE Full text] [doi: 10.1089/g4h.2017.0071] [Medline: 28926286]
23. O'Malley G, Dowdall G, Burls A, Perry I, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR Mhealth Uhealth 2014 Jun 27;2(2):e29 [FREE Full text] [doi: 10.2196/mhealth.3262] [Medline: 25098237]
24. Skinner BF. Verbal Behavior. Eastford, CT: Martino Fine Books; 2015.
25. Caivano S, Ferreira B, Domene S. [Evaluation of the usability of a mobile Digital Food Guide based on user perception]. Cien Saude Colet 2014 May;19(5):1437-1446 [FREE Full text] [doi: 10.1590/1413-81232014195.13932013] [Medline: 24897209]
26. Calinescu A, McLin V, Spahni S, Boggini T, Randriambelonoro M, Jaquier-Grant R, et al. [KidsETransplant: a serious game for liver transplanted children]. Med Sci (Paris) 2016 Dec;32(12):1120-1126 [FREE Full text] [doi: 10.1051/medsci/20163212016] [Medline: 28044976]
27. Dias J, Mekaro M, Cheng Lu JK, Otsuka J, Fonseca L, Zem-Mascarenhas S. Serious game development as a strategy for health promotion and tackling childhood obesity. Rev Lat Am Enfermagem 2016 Aug 15;24:e2759 [FREE Full text] [doi: 10.1590/1518-8345.1015.2759] [Medline: 27533268]
28. Lyles A, Amresh A, Huberty J, Todd M, Lee R. A Mobile, Avatar-Based App for Improving Body Perceptions Among Adolescents: A Pilot Test. JMIR Serious Games 2017 Mar 02;5(1):e4 [FREE Full text] [doi: 10.2196/games.6354] [Medline: 28254737]
29. DeSmet A, Van RD, Compernolle S, Baranowski T, Thompson D, Crombez G. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026]
30. Tark R, Metelitsa M, Akkermann K, Saks K, Mikkel S, Haljas K. Usability, Acceptability, Feasibility, and Effectiveness of a Gamified Mobile Health Intervention (Triumf) for Pediatric Patients: Qualitative Study. JMIR Serious Games 2019 Sep 30;7(3):e13776 [FREE Full text] [doi: 10.2196/13776] [Medline: 31573904]
31. Majumdar D, Koch P, Lee H, Contento I, Islas-Ramos A, Fu D. "Creature-101": A Serious Game to Promote Energy Balance-Related Behaviors Among Middle School Adolescents. Games Health J 2013 Oct;2(5):280-290 [FREE Full text] [doi: 10.1089/g4h.2013.0045] [Medline: 24761326]
32. Martin G, Pear J. Behavior Modification. New York, NY: Pearson; 2010.

Abbreviations

CKD: chronic kidney disease

Edited by G Eysenbach; submitted 26.01.20; peer-reviewed by D Carvalho, S Brigitte; comments to author 02.03.20; revised version received 30.06.20; accepted 11.08.20; published 14.09.20.

Please cite as:
Santana CCDAP, Freitas ATVDS, Oliveira Barreto G, Sousa De Avelar I, Mazaro-Costa R, Bueno GN, Ribeiro DC, Damasceno Silva G, Naghettini AV
Serious Game on a Smartphone for Adolescents Undergoing Hemodialysis: Development and Evaluation
JMIR Serious Games 2020;8(3):e17979
URL: http://games.jmir.org/2020/3/e17979/
doi:10.2196/17979
PMID:32924950

Original Paper

Controlling the Sense of Embodiment for Virtual Avatar Applications: Methods and Empirical Study

Chang-Seop Kim[1], MS; Myeongul Jung[1], BS; So-Yeon Kim[2], PhD; Kwanguk Kim[1], PhD

[1]Department of Computer Science, Hanyang University, Seoul, Republic of Korea

[2]Department of Psychology, Duksung Woman's University, Seoul, Republic of Korea

Corresponding Author:
Kwanguk Kim, PhD
Department of Computer Science
Hanyang University
222 Wangsimni-ro
Seongdong-gu
Seoul, 04763
Republic of Korea
Phone: 82 2 2220 2319
Email: kenny@hanyang.ac.kr

Abstract

Background: The sense of embodiment (SoE) is the feeling of one's own body, and research on the SoE extends from the rubber hand illusion to the full-body ownership illusion with a virtual avatar.

Objective: The key to utilizing a virtual avatar is understanding and controlling the SoE, and it can be extended to several medical applications. In this study, we aimed to clarify these aspects by considering the following three subcomponents of SoE: sense of agency, ownership, and self-location.

Methods: We defined a human avatar (HA), point light avatar (PLA), and out-of-body point light avatar (OBPLA) and compared them in three user studies. In study 1, 28 participants were recruited and the three avatar conditions (HA, PLA, and OBPLA) were compared. In study 2, 29 new participants were recruited, and there were two avatar conditions (HA ad PLA) and two motion synchrony conditions (synchrony and asynchrony). In study 3, 29 other participants were recruited, and there were two avatar conditions (PLA and OBPLA) and two motion synchrony conditions (synchrony and asynchrony). Dependent measures included sense of agency, ownership, and self-location; emotional response; presence; and simulator sickness.

Results: The findings of study 1 showed that the three avatar generation methodologies can control the sense of ownership and self-location in a stepwise manner while maintaining a high sense of agency. In studies 2 and 3, we found dependencies among the three subcomponents of SoE and observed that they affected users' subjective experiences.

Conclusions: Our findings may have implications for boosting the effects of virtual avatar applications in medical areas, by understanding and controlling the SoE with a full-body illusion.

(JMIR Serious Games 2020;8(3):e21879) doi:10.2196/21879

KEYWORDS

embodiment; virtual avatar; full-body illusion; motion capture; virtual reality

Introduction

The sense of embodiment (SoE) is the feeling of one's own body. When we move our bodies in our everyday lives, we feel the SoE unconsciously. The SoE is intimately related to the sense of self and is considered as the starting point of having self-identity [1,2]. The term embodiment has been defined differently in various contexts. From a philosophical perspective, embodiment is considered to be how a person defines and experiences himself/herself [3,4]. In the field of cognitive neuroscience, embodiment is related to how the brain expresses the body [5,6].

Because the SoE is an unconscious process, its alteration or manipulation is difficult. However, Botvinick and Cohen manipulated the SoE using an experiment [7]. With the rubber hand illusion paradigm, the authors demonstrated that participants experienced the SoE on the rubber hand, generating an ownership illusion [7,8]. The rubber hand illusion experiment showed that concurrence of visual and tactile stimulation could generate an embodied experience for a part of one's body. Recently, researchers have begun to apply virtual reality (VR) techniques to expand the SoE to the whole body. A life-sized virtual avatar for a full-body ownership illusion has been developed [9]. Slater et al showed that participants perceived the ownership illusion from virtual arms [10], and Jun et al demonstrated that a full-body ownership illusion and changes in an avatar's facial expression meaningfully affected users' emotions [11].

The effects of the SoE on perceptual and behavioral alterations can be extended to several medical areas. Pioneers have made efforts to apply them in psychological counseling programs [12], and others have used them to enhance cognitive-behavioral therapy for eating disorders [13]. In addition to these applications, systematic desensitized experiences with a virtual avatar could be extended to pain distraction among burn patients, phobias, and posttraumatic stress disorder [14-16]. However, for these purposes, we may need to control the SoE, which has not yet been rigorously examined.

Previous theories have suggested that the SoE can generally be divided into the sense of agency, ownership, and self-location [2]. First, the sense of agency is the subjective feeling that "I am the one who is causing or generating an action" [17]. It includes the subjective experience of action, control, intention, and motor selection, as well as the conscious experience of will [3,18]. Second, the sense of ownership is the feeling that "I am the one who is undergoing the current experience" [17]. The sense of ownership is about the self-attribution of a body that distinguishes it from the sense of agency because it can occur even without a behavior being conducted. Finally, the sense of self-location is the determinate volume in space in which one feels to be located [18]. It is determined by visuospatial perspectives, which are usually egocentric [3], and vestibular signals are also considered to have an important role in self-localization [19].

In this study, we proposed three virtual avatar methods (Figure 1) based on the definition of the three subcomponents of the SoE, as approaches to understand and control the SoE. The first method is human avatar (HA), which is expected to maintain agency, ownership, and self-location at high levels. The HA method uses a human-shaped avatar that moves congruently with the movement of the user and matches the gender and body size of the user [11]. The second method is point light avatar (PLA), which is expected to maintain high levels of agency and self-location but a low level of ownership. The PLA method is based on the point light method proposed by Johansson in his biological motion perception study [20]. The point light is placed at the joint position of the HA to sufficiently reflect the movement of the user but does not have human visual characteristics. Because it does not have human visual characteristics, the PLA is expected to produce a low sense of ownership compared with the HA. However, biological motion is the movement of a living creature and contains information such as behavior, intention, emotion, and personality [21]. Thus, we expect PLA users to be aware of their movement, and their agency will not be lowered. The third method is out-of-body point light avatar (OBPLA), which is expected to maintain a high level of agency and low levels of ownership and self-location. The OBPLA method places the user's viewpoint behind the PLA so that the user can see the PLA from a third person perspective. Therefore, we expect that the OBPLA will show lower self-location compared with the PLA while maintaining a high level of agency. We defined the HA, PLA, and OBPLA and compared them in three user studies. In study 1, we manipulated the SoE by controlling the presence of the virtual avatar and point of view. In studies 2 and 3, we manipulated the SoE by controlling motion synchrony. The differences among the three studies are illustrated in Figure 2.

Figure 1. Proposed virtual avatar methodologies. A: real world; B: human avatar condition; C: point light avatar condition; D: out-of-body point light avatar condition.

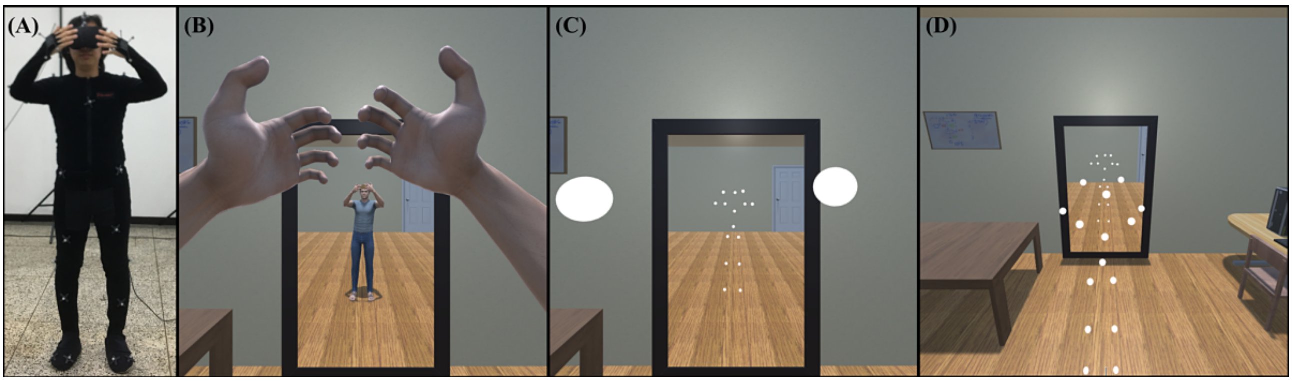
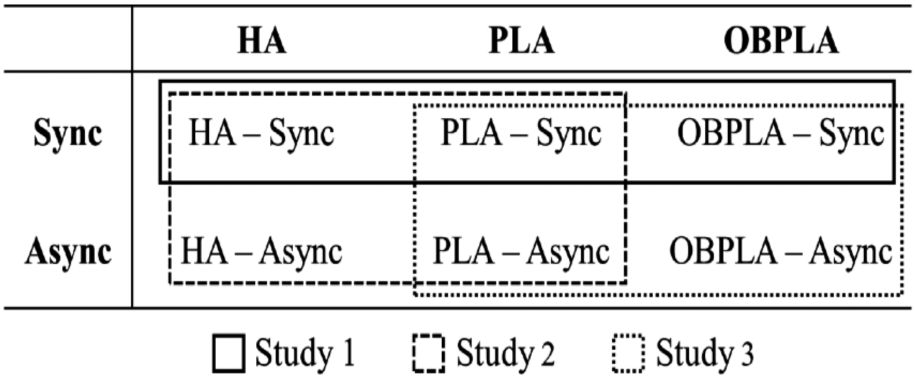


Figure 2. Designs of studies 1, 2, and 3. Async: asynchrony; HA: human avatar; OBPLA: out-of-body point light avatar; PLA: point light avatar; Sync: synchrony.



Methods

Study 1

Participants

Twenty-eight participants (50% female participants, 14/28) were recruited for the experiment (mean age 23.89 years, SD 3.08 years). One participant was excluded from the analysis because of having a high risk of mental illness, as assessed by the Symptom Checklist-90-revised (SCL-90-R; mean score 34.07, SD 40.22) [22]. The experimental protocol was approved by the institutional review board of the research site university.

Hardware and Software Setups

The virtual environment of the experiment was implemented using Unity3D 2017.3.1f (Unity Technologies). The software was executed on a desktop PC with a Nvidia GeForce GTX 1080 graphics card (NVIDIA). The participants wore a head-mounted display (HMD; Oculus Rift CV1, Oculus VR) with a resolution of 1080×1200 pixels per eye and a refresh rate of 90 Hz.

A motion-capture system (Motive 2.0.2; NaturalPoint) was used with 14 Flex13 cameras (NatrualPoint) and 37 markers on a motion-capture suit. Each Flex13 camera had a resolution of 1280×1024 pixels, a frame rate of 120 frames per second (fps), and a latency of 8.33 ms. The positions of the markers within the tracking area were mapped onto a predefined avatar skeleton using Motive 2.0.2. After generating the avatar skeleton, the values for the skeleton bones with six degrees of freedom were streamed using socket communication at 120 fps with a streaming latency of less than 3 ms.

Three Types of Avatars

For this study, the following three avatar conditions were implemented: HA, PLA, and OBPLA.

The HA condition involved moving a human-shaped avatar using a full-body motion capture system. Three Unity3D modules were developed to implement the HA condition. The first module was an avatar animator module, which streamed 21 bones of the skeleton in real time from Motive 2.02 to the avatar. The second module was a size adjuster module, which matched the size of the avatar to the actual participant's body size using the height, shoulder, waist, and pelvis sizes of the participant measured before the experiment. The gender of each avatar was also matched to that of the participant. The third module was a head rotator module, which received the head rotation information of the HMD and rotated the camera in the virtual space, providing visual information to the participant.

The PLA condition involved moving a PLA using a full-body motion capture system. The PLA was composed of 15 spheres, each with a diameter of 5 cm, according to the study of Troje [21]. The 15 spheres were all white without shadows. In the PLA condition, the avatar animator module sent skeleton data to the PLA, and different from the HA condition, the size adjuster module reflected the height, shoulder, and pelvis sizes. The head rotator module was operated the same as under the HA condition. Therefore, the camera in the virtual space rotated and moved to match the head position of the PLA.

The OBPLA condition utilized the out-of-body experience in the PLA condition, so that the PLA could be seen from the outside. The OBPLA condition was created by moving the camera in the PLA condition backward by 120 cm and upward by 30 cm relative to the global coordinate axis. The participants were able to observe their PLAs from a third person perspective.

Virtual Environments

A virtual mirror was placed in the virtual room so that the participants could easily observe their avatar. Participants saw the avatar by both looking directly toward the avatar and by looking at its reflection in the mirror. Several objects, such as computers, a copier, and chairs, were placed in the room for participants to be able to perceive their own body size in relation to these objects.

Dependent Measures

As part of a prequestionnaire, personal information, such as gender and age, was obtained, and the SCL-90-R was used. An embodiment questionnaire (EQ) was used to measure the level of the SoE. A part of the EQ developed by Piryankova et al was selected and modified to fit our experiment [23]. The questions in the EQ are explained in detail in Table 1. The EQ used a seven-point Likert scale to measure agency, ownership, and self-location separately. Participants filled out the EQ after experiencing each condition.

Table 1. List of the items used in the embodiment questionnaire.

Question statement[a]	Sense aspect
I felt I could move the virtual body.	Agency 1
Sometimes I had the feeling that I had control over the virtual body.	Agency 2
I felt as if the virtual body was my body.	Ownership 1
I experienced the arms of virtual body as parts of myself.	Ownership 2
I experienced the legs of virtual body as parts of myself.	Ownership 3
Sometimes I had the feeling that the virtual body belonged to me.	Ownership 4
I experienced the virtual body as myself.	Ownership 5
I felt as if I was inside the virtual body.	Self-location 1
I had the feeling that I was standing in the same location as the virtual body.	Self-location 2

[a]The scale ranges from 1 (fully disagree) to 7 (fully agree).

Procedure

Prior to the experiment, all participants completed consent forms and were instructed on the experimental procedures. All participants also completed a set of prequestionnaires. After the prequestionnaires were completed, participants watched a video clip containing human motion for 5 minutes to prevent participant immobility. Thereafter, participants wore a motion-capture suit and practiced the HA condition for 3 minutes to familiarize themselves with VR and the motion-capture environment. In the main session, participants were asked to move freely in the virtual room under the three conditions. Each condition lasted for 5 minutes and was counterbalanced. After completing each condition, participants removed the HMD and completed the EQ. The purpose of the study was explained to the participants once they had completed all tasks.

Study 2

Study 2 was designed to address the limitations of Study 1. First, we manipulated the level of SoE by controlling motion synchrony. Second, we investigated the differences in subjective measures achieved by controlling the SoE. Specifically, emotion, presence, and simulator sickness were measured. In study 2, there were two avatar conditions (HA and PLA) and two motion synchrony conditions (synchrony and asynchrony). The differences in experimental design between studies 1 and 2 are illustrated in Figure 2.

Participants

Twenty-nine new participants (52% female participants, 15/29) were recruited for study 2. The mean age of the participants was 25.31 years (SD 2.62 years). One of the participants was excluded from the analysis because of the identification of high risk of mental illness according to the SCL-90-R (mean score 39.03, SD 38.41).

Conditions

This study implemented the synchrony and asynchrony conditions to control motion synchrony. The synchrony condition was implemented with the avatar animator, size adjuster, and head rotator modules used in study 1. The asynchrony condition used an avatar animation captured in advance instead of delivering the motion capture data to the avatar through the avatar animator module. Although each participant's movement was not delivered to the avatar, the other conditions were controlled. The size of the avatar was adjusted according to the body size of the participant, and the participant could freely look around even in the asynchrony condition. The avatar conditions (HA and PLA) were the same as those used in study 1.

Dependent Measures

In study 2, personal information and the SCL-90-R were used in the same manner as in study 1. The EQ was the same as in study 1. Furthermore, self-assessment manikin (SAM), presence, and simulator sickness questionnaires were completed. The SAM is a nine-point bipolar scale that uses pictures as a measure of participants' subjective emotional responses to virtual experiences [24]. In this study, arousal and valence were measured as emotional responses. The presence questionnaire (PQ) consists of a seven-point Likert scale and is used to measure presence in VR [25]. The simulator sickness questionnaire (SSQ) is about the degree of simulator sickness and consists of a four-point Likert scale [26].

Study 3

Study 3 was designed to include the OBPLA condition (OBPLA-sync and OBPLA-async). The PLA and OBPLA were used as avatar conditions, and synchrony and asynchrony were used as motion synchrony conditions. This experiment had a 2×2 within-subject design (Figure 2) considering motion synchrony (sync vs async) and avatars (PLA vs OBPLA).

Participants

Twenty-nine participants (48% female participants, 14/29) were recruited for the experiment. The mean age of the participants was 24.14 years (SD 2.74 years). No participants were excluded from the study with the SCL-90-R (mean score 36.86, SD 35.68).

Conditions

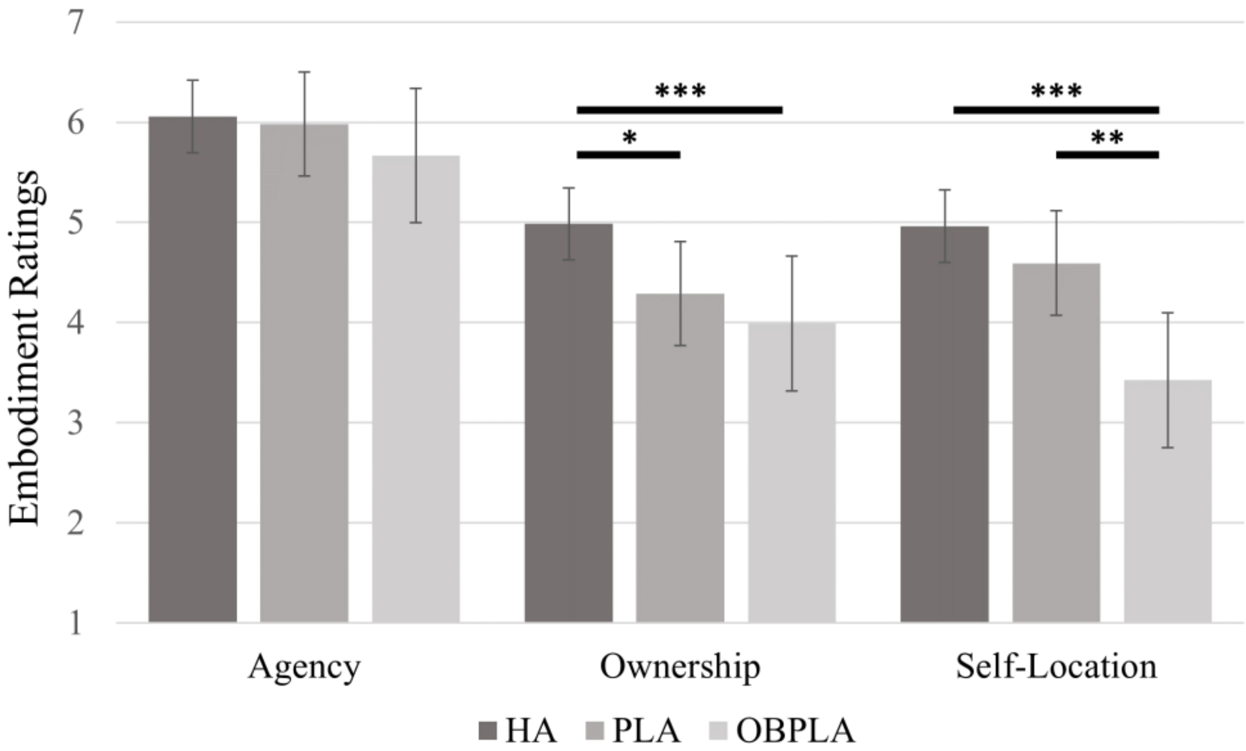
This study used the same synchrony and asynchrony conditions implemented in study 2 to control motion synchrony. The PLA and OBPLA conditions used in study 1 were used as the avatar conditions.

Results

Results of Study 1

We categorized the questions of the EQ in terms of agency, ownership, and self-location. We averaged each category and conducted a repeated measures analysis of variance (ANOVA) to compare the three conditions (HA, PLA, and OBPLA). We performed a *t* test when post-hoc testing was necessary (Figure 3). On comparing the agency scores among the three conditions, there was no statistically significant difference (*P*=.07). On comparing the ownership scores among the three conditions, there was a statistically significant difference ($F_{2,52}$=6.239, *P*=.004, η^2=0.194). In the post-hoc test, the ownership score of the HA was found to be significantly higher than that of the PLA (t_{26}=2.245, *P*=.03) and OBPLA (t_{26}=3.967, *P*<.001). The ownership scores of the PLA and OBPLA conditions showed no statistically significant difference (*P*=.33). On comparing the self-location scores among the three conditions, there was a statistically significant difference ($F_{2,52}$=16.253, *P*<.001, η^2=0.385). The post-hoc *t* test showed that the self-location scores of the HA and PLA conditions were not significantly different (*P*=.11), but the self-location score of the HA condition was significantly higher than that of the OBPLA condition (t_{26}=5.649, *P*<.001). Additionally, the self-location score of the PLA condition was significantly higher than that of the OBPLA condition (t_{26}=3.467, *P*=.002).

Figure 3. Results of the three subcomponents in the embodiment questionnaire for the three conditions. Error bars are one standard deviation. HA: human avatar; OBPLA: out-of-body point light avatar; PLA: point light avatar. **P*<.05, ***P*<.005, ****P*<.001.



Study 1 proposed and compared three avatar conditions. In study 1, the level of agency was maintained under all conditions, and the level of ownership could be controlled by comparing the HA and PLA conditions. Moreover, the level of self-location could be controlled by comparing the PLA and OBPLA conditions. These results suggest that the SoE can be controlled through the proposed methods. Although this study suggested that the SoE can be controlled, there are other factors to consider in further studies. First, changes in agency may have an effect. Second, we must consider additional results that can be achieved by controlling the SoE. Previous studies have shown that subjective factors, such as psychological and emotional effects, may play important roles in virtual avatar application [11,27,28], and those could be affected by controlling the agency, ownership, and self-location.

Results of Study 2

The results of study 2 were analyzed using a 2×2 repeated measures ANOVA with motion synchrony and avatar conditions (Figures 4 and 5). Two factors were analyzed as within-subject factors, and the EQ, SAM, PQ, and SSQ were analyzed. We performed a *t* test when post-hoc testing was necessary.

Figure 4. Box-and-whisker diagrams showing the results of the embodiment questionnaire according to the synchrony and avatar conditions in study 2. Square points are means, thick lines are medians, and round points are outliers. Async: asynchrony; HA: human avatar; PLA: point light avatar; Sync: synchrony.

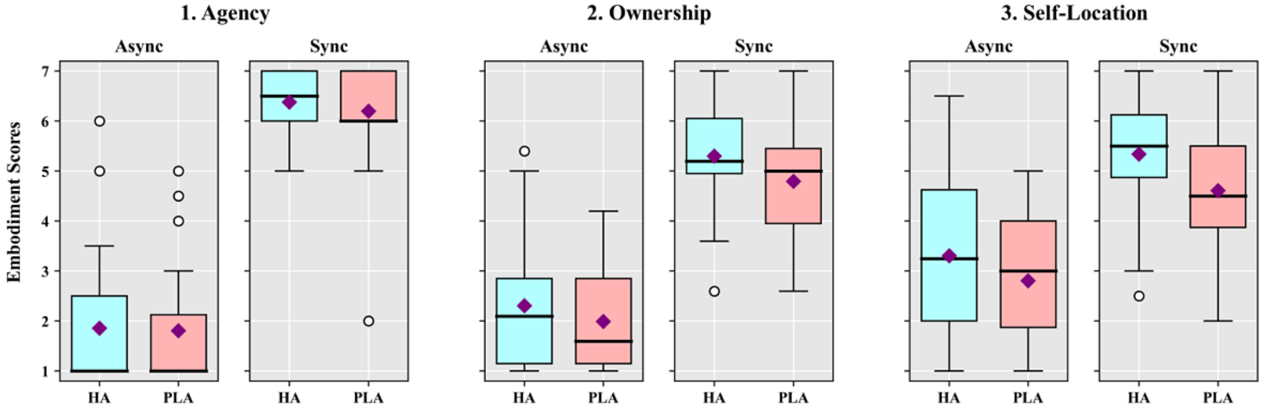
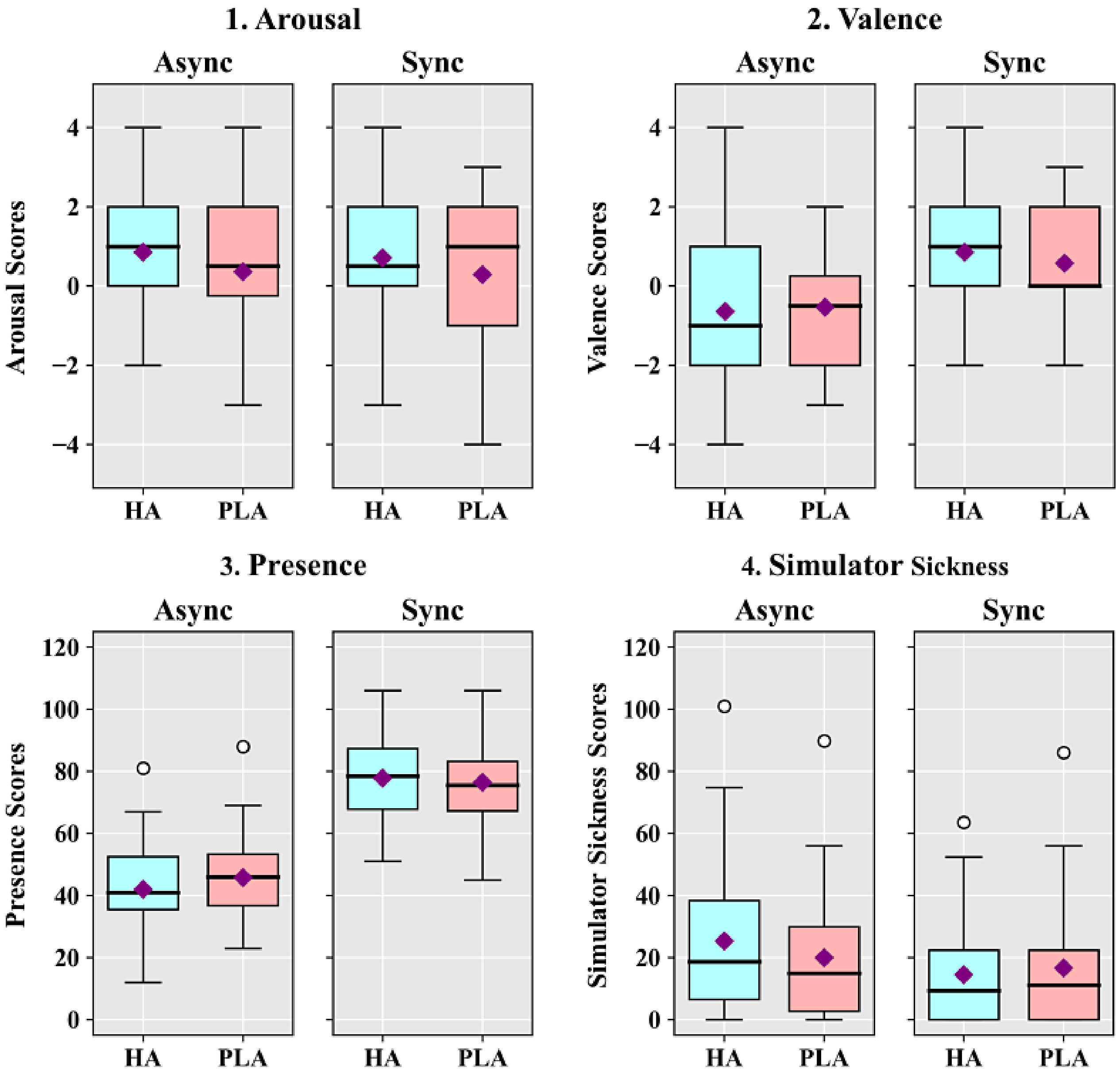


Figure 5. Box-and-whisker diagrams showing the results of the subjective measures according to the synchrony and avatar conditions in study 2. Square points are means, thick lines are medians, and round points are outliers. Async: asynchrony; HA, human avatar; PLA, point light avatar; Sync: asynchrony.



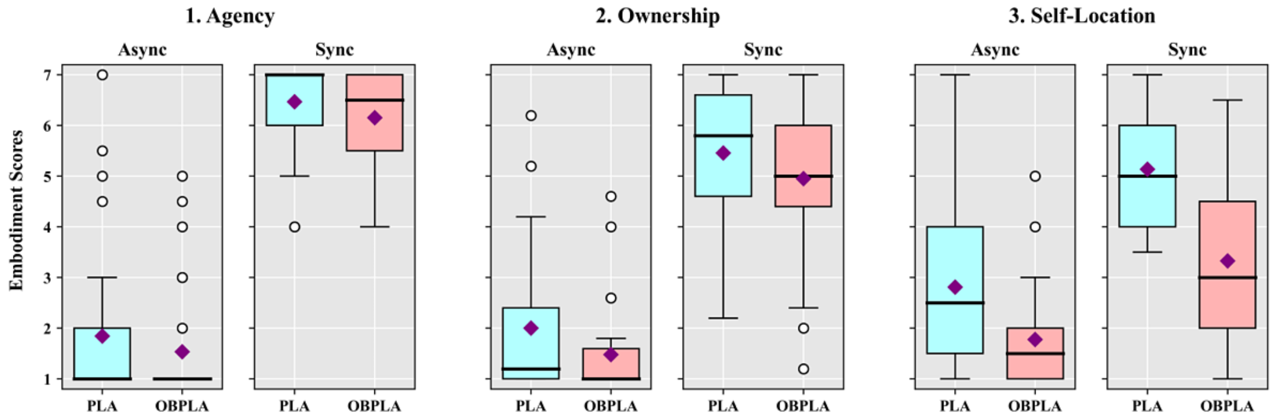
The agency results showed a significant main effect for motion synchrony ($F_{1,27}$=207.328, P<.001, η^2=0.885). There were no significant main effects for avatar (P=.36) or between motion synchrony and avatar (P=.72). The analysis of ownership showed that there were significant main effects for motion synchrony ($F_{1,27}$=102.843, P<.001, η^2=0.792) and avatar ($F_{1,27}$=7.293, P=.01, η^2=0.213). However, there was no significant interaction effect between the two factors (P=.59). The analysis of self-location showed that there were significant main effects for motion synchrony ($F_{1,27}$=37.517, P<.001, η^2=0.582) and avatar ($F_{1,27}$=10.318, P=.003, η^2=0.276). However, there was no significant interaction effect between the two factors (P=.55).

The SAM scores of each condition were analyzed in terms of the differences from baseline measures. For arousal, the results showed no significant main effect for motion synchrony (P=.68), but there was a significant main effect for avatar ($F_{1,27}$=5.725, P=.02, η^2=0.175). There was no significant interaction effect between motion synchrony and avatar (P=.87). The valence analysis showed that there was a significant main effect for motion synchrony ($F_{1,27}$=25.354, P<.001, η^2=0.484). However, there was no significant main effect for avatar (P=.61) and no significant interaction effect between the two factors (P=.23). The PQ results showed a significant main effect for motion synchrony ($F_{1,27}$=85.607, P<.001, η^2=.760), but there was no significant main effect for avatar (P=.29). There was a significant interaction effect between motion synchrony and avatar ($F_{1,27}$=6.063, P=.02, η^2=0.183). Post-hoc *t* tests showed that PQ scores were not significantly different between the HA and PLA under the synchrony condition (P=.43), but the PQ scores under the PLA condition were significantly higher than those under the HA condition in the asynchrony condition (t_{27}=−2.869, P=.008). The SSQ analysis showed that there was a significant main effect for motion synchrony ($F_{1,27}$=4.784, P=.04, η^2=0.151). There was no significant main effect for avatar (P=.46) and no significant interaction effect between the two factors (P=.12).

Results of Study 3

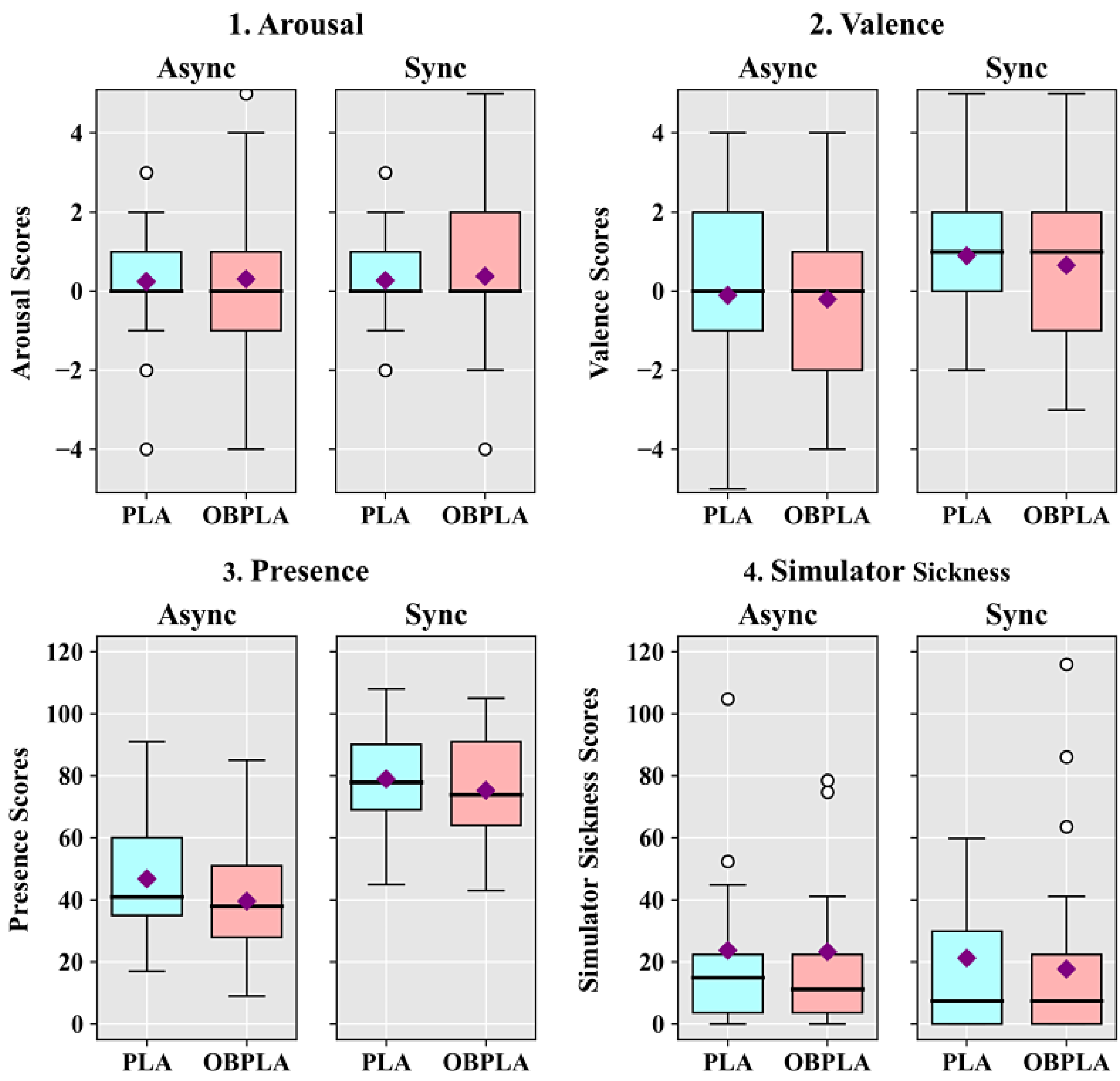
The results of study 3 were analyzed using a 2×2 repeated measures ANOVA with the motion synchrony and avatar conditions (Figures 6 and 7).

Figure 6. Box-and-whisker diagrams showing the results of the embodiment questionnaire according to the synchrony and avatar conditions in study 3. Square point are means, thick lines are medians, and round points are outliers. Async: asynchrony; OBPLA: out-of-body point light avatar; PLA: point light avatar; Sync: asynchrony.



The agency results showed significant main effects for motion synchrony ($F_{1,28}$=218.892, P<.001, η^2=0.887) and avatar ($F_{1,28}$=6.610, P=.02, η^2=0.191). However, there was no significant interaction effect between motion synchrony and avatar (P=1.00). The analysis of ownership showed significant main effects for motion synchrony ($F_{1,28}$=122.502, P<.001, η^2=0.814) and avatar ($F_{1,28}$=14.283, P<.001, η^2=0.338). However, there was no significant interaction effect between the two factors (P=.97). For self-location, the results showed significant main effects for motion synchrony ($F_{1,28}$=40.088, P<.001, η^2=0.589) and avatar ($F_{1,28}$=40.180, P<.001, η^2=0.589). There was a significant interaction effect between motion synchrony and avatar ($F_{1,28}$=5.156, P=.03, η^2=0.156). In the post-hoc test, the self-location score of the PLA was significantly higher than that of the OBPLA under the synchrony condition (t_{28}=7.007, P<.001). Under the asynchrony condition, the self-location score of the PLA was also higher than that of the OBPLA (t_{28}=3.405, P=.002), but the difference was greater under the synchrony condition.

Figure 7. Box-and-whisker diagrams showing the results of the subjective measures according to the synchrony and avatar conditions in study 3. Square points are means, thick lines are medians, and round points are outliers. Async: asynchrony; OBPLA: out-of-body point light avatar; PLA: point light avatar; Sync: synchrony.



The arousal results showed no significant main effects for motion synchrony (P=.75) and avatar (P=.56). There was no significant interaction effect between motion synchrony and avatar (P=.94). The valence results showed that there was a significant main effect for motion synchrony ($F_{1,28}$=20.190, P<.001, η^2=0.419). However, there was no significant main effect for avatar (P=.29) and no significant interaction effect between the two factors (P=.76). The PQ results showed a significant main effect for motion synchrony ($F_{1,28}$=87.779, P<.001, η^2=0.758) and a significant main effect for avatar ($F_{1,28}$=6.881, P=.01, η^2=0.197). However, there was no significant interaction effect between motion synchrony and avatar (P=.26). The analysis of the SSQ showed that there was a significant main effect for motion synchrony ($F_{1,28}$=7.095, P=.01, η^2=0.202). However, there was no significant main effect for avatar (P=.46) and no significant interaction effect between the two factors (P=.47).

Discussion

In this research, we proposed virtual avatar methodologies to understand and control the SoE and compared the HA, PLA, and OBPLA conditions using full-body motion capture and VR technologies. Study 1 suggested that it is possible to control the level of SoE from the HA to PLA conditions and from the PLA to OBPLA conditions. In studies 2 and 3, we manipulated motion synchrony and avatar type to determine whether such manipulations affect the SoE, emotional response, presence, and simulator sickness. The results suggested that motion synchrony affects all subtypes of SoE. Regarding avatar type, differences in both ownership and self-location were found between the HA and PLA conditions, indicating positive effects of the HA compared with the PLA for those two SoE variables. Additionally, we found that the PLA increased agency, ownership, and self-location compared with the OBPLA. Presence, emotion, and simulator sickness were affected by motion synchrony and avatar conditions. Our findings suggest that the SoE and subjective experiences can be affected differently with different avatar types in the full-body illusion. We believe that these results can be applied to various fields. For example, the results can be used for systematic desensitization of phobias or pain reduction by controlling the level of SoE. They can also be used to control the effects of emotion and presence associated with using virtual avatars.

In study 1, no relevant differences were found for agency among the three conditions. Despite the absence of human visual characteristics, if full-body motion-capture is maintained, the level of agency is not lowered. Recently, Zopf et al [29] compared virtual fingers and a virtual sphere moving correspondingly to a participant's hand and found that the moving sphere did not lower the level of agency. This study suggests that the degree of visual stimulation does not affect the sense of agency for not only the hand but also the whole body. Therefore, based on existing research and this study, we suggest that movements induced by one's intent and predictions determine the sense of agency. Moreover, in the comparison among the avatar conditions, we found that the details of the avatar's visual characteristics did not affect the sense of agency.

In studies 2 and 3, we investigated the effects of motion synchrony on sense of agency, ownership, and self-location, and the results suggested that motion synchrony greatly affects all SoE subcomponents. Interestingly, the effects of motion synchrony overwhelmed all other effects. Therefore, motion synchrony may need to be controlled if we wish to control the SoE in general. However, the results of study 1 also suggested that the sense of body ownership and self-location can be controlled once motion synchrony is satisfied. Merging the results from studies 1, 2, and 3, we expect that we can use motion synchrony for the general control of the SoE, and we can use the presence of the virtual body and point of view for more detailed control over the SoE. The sense of ownership under the HA condition was higher than that under the PLA and OBPLA conditions. This result is an extension of previous findings that the sense of ownership is induced when the virtual body and physical body are morphologically coincident [7,30,31]. This also suggests that the HA and PLA are appropriate tools for controlling the level of ownership. The point light (PL) method is known to have sufficient information to not only express human behavior but also determine gender [32,33] and recognize emotion [34]. On the other hand, human visual features, such as important interjoint expressions, finger expressions, rotation of some joints, and skin texture are excluded in the PL method. These characteristics of the PL method seem to be suitable for controlling ownership when the PLA is used as a virtual body. Participants experienced a higher sense of self-location under the HA and PLA conditions than under the OBPLA condition. This result supports previous findings that self-location can be controlled through viewpoint changes in low agency situations [35,36]. This study suggested that the control of self-location through changes of perspective is also possible in a high agency situation, such as in a full-body motion-capture system. In study 1, there was no relevant difference in the sense of self-location between the HA and PLA conditions, but in study 2, there was a difference in the sense of self-location according to the avatar's visual characteristics. Moreover, in study 3, we found that changes in the viewpoint (first vs third person perspective) affected the sense of agency. This suggests the possibility of mutual dependencies among the three subcomponents or the possibility of the context effect of synchrony and asynchrony. However, additional studies are required on this issue.

This study also found that there were significant differences in emotional response, presence, and simulator sickness according to avatar type and motion synchrony. Although the design of this study did not include an emotional evocation context, it was shown that the participants felt different emotional levels depending on the condition. Participants felt higher arousal under the HA condition than under the PLA condition. The HA expresses a human face, but the PLA expresses the face as point light, so the HA contains more facial information than the PLA. Although the avatar under the HA condition had a neutral face, it is possible that the human appearance influenced arousal. The present results also showed an interaction effect between motion synchrony and the avatar's visual characteristics and a main effect in the avatar's perspective. It would be beneficial to control presence more precisely by using these combinations of SoE subcomponents.

This study has some limitations. First, the participants in our study were all healthy young college students. It is possible that the characteristics of the participants affected the results of embodiment and subjective responses. Therefore, in future studies, it is necessary to collect data from participants of various ages and health statuses. Second, this study asked participants to move freely and measured arousal and valence as emotions, but pain, anxiety, and fear were not accurately represented. These emotions should be considered for future medical applications. Therefore, in future studies, it is necessary to give specific tasks to participants and check how these methods of controlling the embodiment affect participants in emotional tasks. Third, this study proposed the HA, PLA, and OBPLA to control the external appearance and perspective of the avatar. However, an extension of these methods can be used to control the SoE. For example, we can use lines to express the body or low-polygon human avatars to control the level of visual characteristics. We can also control the point of view, allowing the human avatar to be seen from another third person perspective, or we can see our avatar from the front. Future studies must use additional methods to regulate the SoE and identify the impacts.

In this study, we proposed and compared the HA, PLA, and OBPLA conditions to understand and control the SoE. We also studied how emotional response, presence, and simulator sickness can be affected by controlling the subcomponents of the SoE. The results suggested that the three avatar generation methodologies and two synchrony levels can control the sense of agency, ownership, and self-location, and emotional response, presence, and simulator sickness were differently affected by each methodology. This study may have implications for boosting the effects of virtual avatar applications in several fields of study, such as psychotherapy, entertainment, and education, by controlling the SoE with a full-body illusion.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C7A1029688 and NRF-2018R1A5A7059549). The funders were not involved in the design or conduct of the study; data collection; management; analysis; interpretation; or preparation, review, or approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Cassam Q. Self and World. Oxford, UK: Oxford University Press; 1997.
2. Longo MR, Schüür F, Kammers MP, Tsakiris M, Haggard P. What is embodiment? A psychometric approach. Cognition 2008 Jun;107(3):978-998. [doi: 10.1016/j.cognition.2007.12.004] [Medline: 18262508]
3. Blanke O, Metzinger T. Full-body illusions and minimal phenomenal selfhood. Trends Cogn Sci 2009 Jan;13(1):7-13. [doi: 10.1016/j.tics.2008.10.003] [Medline: 19058991]
4. Metzinger T. Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples. Prog Brain Res 2008;168:215-245. [doi: 10.1016/S0079-6123(07)68018-2] [Medline: 18166398]
5. Berlucchi G, Aglioti S. The body in the brain: neural bases of corporeal awareness. Trends Neurosci 1997 Dec;20(12):560-564. [doi: 10.1016/s0166-2236(97)01136-3] [Medline: 9416668]
6. Graziano MS, Botvinick MM. How the brain represents the body: insights from neurophysiology and psychology. In: Prinz W, Hommel B, editors. Common Mechanisms in Perception and Action: Attention and Performance. Oxford, UK: Oxford University Press; 2002:136-157.
7. Botvinick M, Cohen J. Rubber hands 'feel' touch that eyes see. Nature 1998 Feb 19;391(6669):756. [doi: 10.1038/35784] [Medline: 9486643]
8. Armel KC, Ramachandran VS. Projecting sensations to external objects: evidence from skin conductance response. Proc Biol Sci 2003 Jul 22;270(1523):1499-1506 [FREE Full text] [doi: 10.1098/rspb.2003.2364] [Medline: 12965016]
9. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. PLoS One 2010 May 12;5(5):e10564 [FREE Full text] [doi: 10.1371/journal.pone.0010564] [Medline: 20485681]
10. Slater M, Perez-Marcos D, Ehrsson HH, Sanchez-Vives MV. Inducing illusory ownership of a virtual body. Front Neurosci 2009 Sep;3(2):214-220 [FREE Full text] [doi: 10.3389/neuro.01.029.2009] [Medline: 20011144]
11. Jun J, Jung M, Kim SY, Kim K. Full-Body Ownership Illusion Can Change Our Emotion. In: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada. [doi: 10.1145/3173574.3174175]
12. Osimo SA, Pizarro R, Spanlang B, Slater M. Conversations between self and self as Sigmund Freud--A virtual body ownership paradigm for self counselling. Sci Rep 2015 Sep 10;5:13899 [FREE Full text] [doi: 10.1038/srep13899] [Medline: 26354311]
13. Cesa GL, Manzoni GM, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality for enhancing the cognitive behavioral treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. J Med Internet Res 2013;15(6):e113 [FREE Full text] [doi: 10.2196/jmir.2441] [Medline: 23759286]
14. Hoffman HG, Patterson DR, Carrougher GJ. Use of virtual reality for adjunctive treatment of adult burn pain during physical therapy: a controlled study. Clin J Pain 2000 Sep;16(3):244-250. [doi: 10.1097/00002508-200009000-00010] [Medline: 11014398]
15. Reger GM, Gahm GA. Virtual reality exposure therapy for active duty soldiers. J Clin Psychol 2008 Aug;64(8):940-946. [doi: 10.1002/jclp.20512] [Medline: 18612993]
16. Tardif N, Therrien C, Bouchard S. Re-Examining Psychological Mechanisms Underlying Virtual Reality-Based Exposure for Spider Phobia. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):39-45. [doi: 10.1089/cyber.2017.0711] [Medline: 30256675]
17. Gallagher S. Philosophical conceptions of the self: implications for cognitive science. Trends in Cognitive Sciences 2000 Jan;4(1):14-21. [doi: 10.1016/s1364-6613(99)01417-5] [Medline: 10637618]
18. Kilteni K, Groten R, Slater M. The Sense of Embodiment in Virtual Reality. Presence: Teleoperators and Virtual Environments 2012 Nov;21(4):373-387. [doi: 10.1162/pres_a_00124]
19. Lopez C, Halje P, Blanke O. Body ownership and embodiment: vestibular and multisensory mechanisms. Neurophysiol Clin 2008 Jun;38(3):149-161. [doi: 10.1016/j.neucli.2007.12.006] [Medline: 18539248]
20. Johansson G. Visual perception of biological motion and a model for its analysis. Perception & Psychophysics 1973 Jun;14(2):201-211. [doi: 10.3758/bf03212378]

21. Troje NF. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. J Vis 2002;2(5):371-387 [FREE Full text] [doi: 10.1167/2.5.2] [Medline: 12678652]
22. Derogatis LR. SCL-90-R: Administration, scoring & procedures manual-II for the (revised) version and other instruments of the psychopathology rating scale series. Baltimore, MD: Clinical Psychometric Research; 1992.
23. Piryankova IV, Wong HY, Linkenauger SA, Stinson C, Longo MR, Bülthoff HH, et al. Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. PLoS One 2014 Aug 1;9(8):e103428 [FREE Full text] [doi: 10.1371/journal.pone.0103428] [Medline: 25083784]
24. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
25. Witmer BG, Singer MJ. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
26. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
27. Kim S, Park H, Jung M, Kim KK. Impact of Body Size Match to an Avatar on the Body Ownership Illusion and User's Subjective Experience. Cyberpsychol Behav Soc Netw 2020 Apr;23(4):234-241. [doi: 10.1089/cyber.2019.0136] [Medline: 32074457]
28. Yoshie M, Haggard P. Negative emotional outcomes attenuate sense of agency over voluntary actions. Curr Biol 2013 Oct 21;23(20):2028-2032 [FREE Full text] [doi: 10.1016/j.cub.2013.08.034] [Medline: 24094850]
29. Zopf R, Polito V, Moore J. Revisiting the link between body and agency: visual movement congruency enhances intentional binding but is not body-specific. Sci Rep 2018 Jan 09;8(1):196 [FREE Full text] [doi: 10.1038/s41598-017-18492-7] [Medline: 29317726]
30. IJsselsteijn WA, de Kort YA, Haans A. Is This Hand I See Before Me? The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality. Presence: Teleoperators and Virtual Environments 2006 Aug 01;15(4):455-464. [doi: 10.1162/pres.15.4.455]
31. Sanchez-Vives MV, Spanlang B, Frisoli A, Bergamasco M, Slater M. Virtual hand illusion induced by visuomotor correlations. PLoS One 2010 Apr 29;5(4):e10381 [FREE Full text] [doi: 10.1371/journal.pone.0010381] [Medline: 20454463]
32. Barclay CD, Cutting JE, Kozlowski LT. Temporal and spatial factors in gait perception that influence gender recognition. Percept Psychophys 1978 Feb;23(2):145-152. [doi: 10.3758/bf03208295] [Medline: 643509]
33. Kozlowski LT, Cutting JE. Recognizing the sex of a walker from a dynamic point-light display. Perception & Psychophysics 1977 Nov;21(6):575-580. [doi: 10.3758/bf03198740]
34. Dittrich WH, Troscianko T, Lea SE, Morgan D. Perception of emotion from dynamic point-light displays represented in dance. Perception 1996;25(6):727-738. [doi: 10.1068/p250727] [Medline: 8888304]
35. Lenggenhager B, Smith ST, Blanke O. Functional and neural mechanisms of embodiment: importance of the vestibular system and the temporal parietal junction. Rev Neurosci 2006;17(6):643-657. [doi: 10.1515/revneuro.2006.17.6.643] [Medline: 17283609]
36. Maselli A, Slater M. The building blocks of the full body ownership illusion. Front Hum Neurosci 2013;7:83 [FREE Full text] [doi: 10.3389/fnhum.2013.00083] [Medline: 23519597]

Abbreviations

ANOVA: analysis of variance
EQ: embodiment questionnaire
fps: frames per second
HA: human avatar
HMD: head-mounted display
OBPLA: out-of-body point light avatar
PL: point light
PLA: point light avatar
PQ: presence questionnaire
SAM: self-assessment manikin
SCL-90-R: Symptom Checklist-90-revised
SoE: sense of embodiment
SSQ: simulator sickness questionnaire
VR: virtual reality

Edited by G Eysenbach; submitted 27.06.20; peer-reviewed by SH Choi, J Kim; comments to author 04.08.20; revised version received 18.08.20; accepted 18.08.20; published 22.09.20.

Please cite as:
Kim CS, Jung M, Kim SY, Kim K
Controlling the Sense of Embodiment for Virtual Avatar Applications: Methods and Empirical Study
JMIR Serious Games 2020;8(3):e21879
URL: http://games.jmir.org/2020/3/e21879/
doi:10.2196/21879
PMID:32960174

Original Paper

Lessons Learned From an Evaluation of Serious Gaming as an Alternative to Mannequin-Based Simulation Technology: Randomized Controlled Trial

Natalie C Benda[1,2], PhD; Kathryn M Kellogg[3,4], MD, MPH; Daniel J Hoffman[1], BS; Rollin J Fairbanks[3,4], MD, MS; Tamika Auguste[5,6,7], MD

[1]National Center for Human Factors in Healthcare, MedStar Institute for Innovation, MedStar Health, Washington, DC, United States
[2]Division of Health Informatics, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States
[3]Emergency Medicine, Georgetown University School of Medicine, Washington, DC, United States
[4]Quality and Safety, MedStar Health, Columbia, MD, United States
[5]Women's and Infants' Services, MedStar Washington Hospital Center, Washington, DC, United States
[6]Obstetrics and Gynecology, Georgetown University School of Medicine, Washington, DC, United States
[7]Simulation Training & Education Lab (SiTEL), MedStar Health, Washington, DC, United States

Corresponding Author:
Tamika Auguste, MD
Women's and Infants' Services
MedStar Washington Hospital Center
110 Irving St. NW
5B-41
Washington, DC, 20010
United States
Phone: 1 202 877 8177
Email: tamika.c.auguste@medstar.net

Abstract

Background: The use of new technology like virtual reality, e-learning, and serious gaming can offer novel, more accessible options that have been demonstrated to improve learning outcomes.

Objective: The aim of this study was to compare the educational effectiveness of serious game–based simulation training to traditional mannequin-based simulation training and to determine the perceptions of physicians and nurses. We used an obstetric use case, namely electronic fetal monitoring interpretation and decision making, for our assessment.

Methods: This study utilized a mixed methods approach to evaluate the effectiveness of the new, serious game–based training method and assess participants' perceptions of the training. Participants were randomized to traditional simulation training in a center with mannequins or serious game training. They then participated in an obstetrical in-situ simulation scenario to assess their learning. Participants also completed a posttraining perceptions questionnaire.

Results: The primary outcome measure for this study was the participants' performance in an in-situ mannequin-based simulation scenario, which occurred posttraining following a washout period. No significant statistical differences were detected between the mannequin-based and serious game–based groups in overall performance, although the study was not sufficiently powered to conclude noninferiority. The survey questions were tested for significant differences in participant perceptions of the educational method, but none were found. Qualitative participant feedback revealed important areas for improvement, with a focus on game realism.

Conclusions: The serious game training tool developed has potential utility in providing education to those without access to large simulation centers; however, further validation is needed to demonstrate if this tool is as effective as mannequin-based simulation.

(JMIR Serious Games 2020;8(3):e21123) doi:10.2196/21123

KEYWORDS

simulation training; continuing medical education; obstetrics

Introduction

Continuing education and maintenance of competency of health care professionals are critical to patient safety [1,2]. Clinical simulation training has been proven to teach and refine skills, offering the realism of the clinical environment without risks to patient safety [1,2]. However, many barriers exist to the universal availability of simulation, and an alternative simulation technology is needed. The current standard in simulation training is mannequin-based, human-patient simulators (example in Figure 1). Although its effective use has expanded rapidly in the last decade, this training modality requires expensive equipment, specialized instructors, and ongoing infrastructure support [3-5]. Widespread implementation throughout the United States has been impeded by the high cost and shortage of simulation training expertise outside of academic and large multihospital medical centers. Even in areas that do have simulation centers, few practicing physicians participate regularly [3,6].

Figure 1. NOELLE S550 Maternal and Neonatal Birthing Simulator depiction.

Serious games offer a viable alternative to mannequin-based simulation. Serious games provide a learner-oriented approach, in which the learner can control the entire learning process [7,8]. This approach has also been demonstrated to have lower overall costs than mannequin-based simulation programs [8]. The use of serious games has shown promise for effectiveness in learning outcomes as compared to conventional training methods (eg, in-person or online lectures, didactic case study discussion) related to, for example, diabetes care (with primary care providers), surgical training, and emergency airway management [9-14]. However, few studies have evaluated serious games as an alternative to mannequin-based, human-patient simulators [15]. In 4 recent review articles regarding the efficacy of serious games for training health care professionals, only 1 article compared serious games to mannequin-based simulation [16-19]. Cendan and Johnson [15] compared the use of serious game and mannequin-based simulation for teaching shock physiology to second-year medical students. The authors did not detect significant differences between the serious game and mannequin-based simulation conditions in knowledge related to cardiac shock physiology and treatment among medical students. However, this study was not sufficiently powered to conclude noninferiority of the 2 treatments, and students significantly preferred mannequin-based simulation [15].

None of the 4 aforementioned review articles included serious game applications for obstetrics [16-19]. Obstetrics is one of the highest risk areas in health care [20], making it an ideal area for trialing new, innovative training techniques. Three-quarters of US obstetrician gynecologists (OBGYNs) will face a litigation claim by the age of 45 years [20]. Electronic fetal monitoring (EFM) is considered the standard of care for OBGYNs to monitor the status of the fetus during labor. Skills in EFM interpretation (determining baseline heart rate, accelerations in heart rate, decelerations in heart rate, and how this represents the current status of the fetus) and the knowledge of how to apply validated treatment protocols are critical to safe deliveries. Fetal monitoring skills are highly variable among practitioners and are difficult to teach, but clinical (mannequin-based) simulation has been proven to impart critical skills that improve patient outcomes by reducing errors and delays in care [21-25]. The importance of teaching EFM skills, coupled with the difficulties in providing widespread, accessible mannequin-based simulation to obstetrics providers and nurses, make this an excellent application area for testing the viability of serious games.

In order to address this important problem, our team of experts in education, clinical simulation, human factors engineering, obstetrics, and serious game development collaborated to create a serious game–based simulation for obstetrical training.

The aim of this study was to compare the educational effectiveness of serious game–based simulation training to traditional mannequin-based simulation training and to determine the perceptions of providers and nurses in their experience using serious game–based simulation. Our hypothesis was that the serious game–based training would be noninferior to the mannequin-based training in terms of educational effectiveness. We used EFM interpretation and decision-making skills as our test case. We also assessed participant perceptions through both quantitative and qualitative feedback to provide actionable results.

Methods

Setting and Population

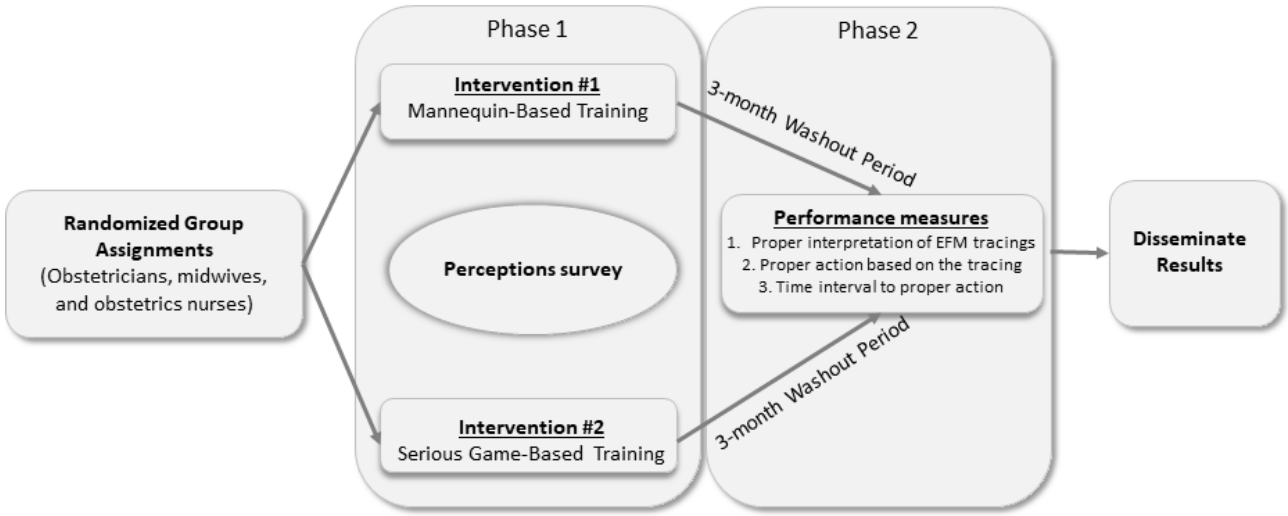
This study was conducted across 7 diverse hospitals with obstetrical services within a not-for-profit health care system in the mid-Atlantic region of the United States. The hospitals included academic medical centers and community, urban, and suburban hospitals of varying size. The health care system averages approximately 12,000 deliveries a year.

The participants recruited for this study were attending and resident OBGYNs, midwives, and perinatal nurses. All recruited participants worked in the health care system's labor and deliveries or mother baby units between July 2012 and November 2015. Those that had participated in beta testing of the serious game or those who would not be able to complete both phases of the study were excluded.

Study Design

This study utilized a mixed methods approach to evaluate the effectiveness of the new, serious game–based training method and to assess participants' perceptions of the training. A randomized controlled trial was performed to compare the educational effectiveness of the serious game–based virtual simulation to traditional mannequin-based simulation. This study was conducted in 2 phases. In Phase 1, participants completed a verbal informed consent process and were randomized into the mannequin-based training group (Mannequin Group) or the serious game–based training group (Game Group). Upon completion of the Phase 1 training session, participants completed a questionnaire to assess their perceptions of the training. Following at least a 3-month washout period, both treatment groups participated in a posttest performance assessment. A 3-month washout period was chosen based on availability of the participants, availability of the simulation center, and the already scheduled in-situ drills that were the posttest performance assessments. This design is depicted in Figure 2. The health system's institutional review board approved this study.

Figure 2. Study design flow diagram. EFM: electronic fetal monitoring.



Simulation Scenarios and Interventions

Simulation Scenarios

The scenarios used in both treatment groups were developed through collaboration between the subject matter experts in obstetrics and Certified Healthcare Simulation Educators at the Simulation Training & Education Lab of MedStar Health. Scenarios were designed to be representative of true clinical scenarios with clear decision-making points. They were based on simulation scenarios previously developed and used by our group for postgraduate education, postgraduate remediation, and in-situ team training drills. In each intervention group, the participant was expected to manage the simulated laboring patient in a realistic clinical scenario from triage to delivery. Time elapsed to cover 4 discrete scenarios that involved interpreting the EFM strips that were correlated to the patient's worsening preeclampsia and fetal distress. The scenarios were delivered through 2 different interventions or mediums (described in the following sections): mannequin-based simulation training (Mannequin Group) or serious game–based simulation training (Game Group).

Mannequin-Based Training

Participants randomized to the Mannequin Group completed the scenario at a Simulation Training and Education Lab simulation center site involving a pregnant mannequin patient (NOELLE S550 Maternal and Neonatal Birthing Simulator, Gaumard, Miami, FL; Figure 1) with preeclampsia with severe features. Participants interacted with the mannequin as they would with a typical patient. The training was scripted into 4 discrete scenes of the scenario with a single participant being run at a time. The facilitator was 1 of 2 experienced obstetricians with at least 7 years of facilitating training for residents in simulation labs with interdisciplinary teams in in-situ drills. Pilot testing of the completed scenarios and simulation was done with a group of physicians and nurses. The training was done once, and no repetitions were permitted.

Serious Game–Based Training

Participants randomized to the Game Group completed the virtual simulation session on their personal computer. This simulation program runs on a standard personal computer; thus, it can be easily used in the hospital, at home, or anywhere a computer and internet connection are available. The EFM serious game was developed utilizing the MedStar Digital Simulation Platform and the Unity3D game engine. The game also went through extensive pilot and usability testing after development, with those potential learners being excluded from participating in the subsequent study. The feedback that was given in the pilot testing was used to improve the final serious game product. Most learners accessed the EFM Trainer via a personal computer using web browsers via a learning management system. Participants completed the training at their leisure; therefore, any external stimuli varied. The training was done once, and no repetitions were permitted.

The "avatar" in the serious game represented the participant (the provider or nurse using the system). The participant's avatar performed real-life tasks in a realistic patient scenario. This is depicted in Figure 3. The scenarios in the serious game were clinically the same as the scenarios used in the Mannequin Group. Any difference noted was slight and due to any difficulty representing the nuance in a serious game. Items like presenting symptoms, differential diagnoses, exams, laboratory values, and EFM tracings were the same. Names of patients and visual representation of the patients differed only slightly.

Figure 3. Depiction of the avatar participants utilized to navigate the serious game.

Measurements

Posttest Performance Assessment

The Phase 2 posttest assessment involved participation in the MedStar Obstetrical Safety Training (MOST) program, a mannequin-based simulation program that served as the gold standard for performance assessment. The hospital system implemented the MOST program 10 years ago, and all practicing obstetrics physicians and nurses are required to complete it. The MOST program is a single-session, in-situ, mannequin-based simulation training program that allows perinatal teams to practice and assess various aspects of emergent obstetrical care, including interpretation and decision making related to EFM strips. For the purposes of this study, the MOST program session was used as the postintervention tool to assess the learner's knowledge after the minimum 3-month washout period following the initial trainings described (Game Group or Mannequin Group), not as an educational program.

A newly developed evaluation tool was incorporated into the MOST training and was used to assess performance on standardized scenarios involving the need for EFM interpretation and related decision making. Using the MOST program for the evaluation was chosen because it closely approximates real-life performance. To avoid any scenario confounds, the specific scenarios selected for MOST testing were different from those presented to either group during training. The slight differences were based around the types of categories of EFM tracings and the time it took for abnormal tracings to develop.

Posttraining Perceptions Questionnaire

We developed a survey instrument in conjunction with a PhD-level biostatistician to assess perceptions of the intended user groups, administered at completion of Phase 1. Prior to the randomized controlled trial, the survey was pilot tested by those participating in the pilot testing for the serious game to ensure comprehensibility. The versions of the survey administered to each group differed only in wording to be relevant to the arm of the study (ie, either referred to serious game–based simulation or mannequin-based simulation). The questions regarding the participants' experience consisted of 19 statements pertaining to the participants' experience with technology, views on simulation, and their specific experience with this study. The participants rated these statements on a 5-point Likert scale with responses ranging from strongly disagree to strongly agree. At the conclusion of the questionnaire, participants also provided free-text narratives regarding their experience.

Study Protocol and Data Collection

Prestudy: Recruitment and Randomization

Recruitment occurred via email, dissemination of paper fliers, principle investigator attendance of staff meetings, and word-of-mouth. Interested parties contacted research personnel (NB or DH) who ensured prospective participants met the inclusion criteria and obtained verbal consent. Eligible participants were then randomized to the Mannequin Group or Game Group. A randomized permuted block design of mixed block size was used to assign participants to 1 of the 2 study groups with a 1:1 ratio. The randomization sequence was generated by a PhD-level biostatistician, and allocation was performed by a trained study research assistant.

Phase 1: Simulation Training

Participants randomized to the mannequin-based training were first briefed by a clinical simulation specialist regarding the functionality of the mannequin and how to verbalize their interpretation of the EFM strip and subsequent actions. There was also a visible, written sign reminding the participants which components of their EFM strip interpretation needed to be verbalized.

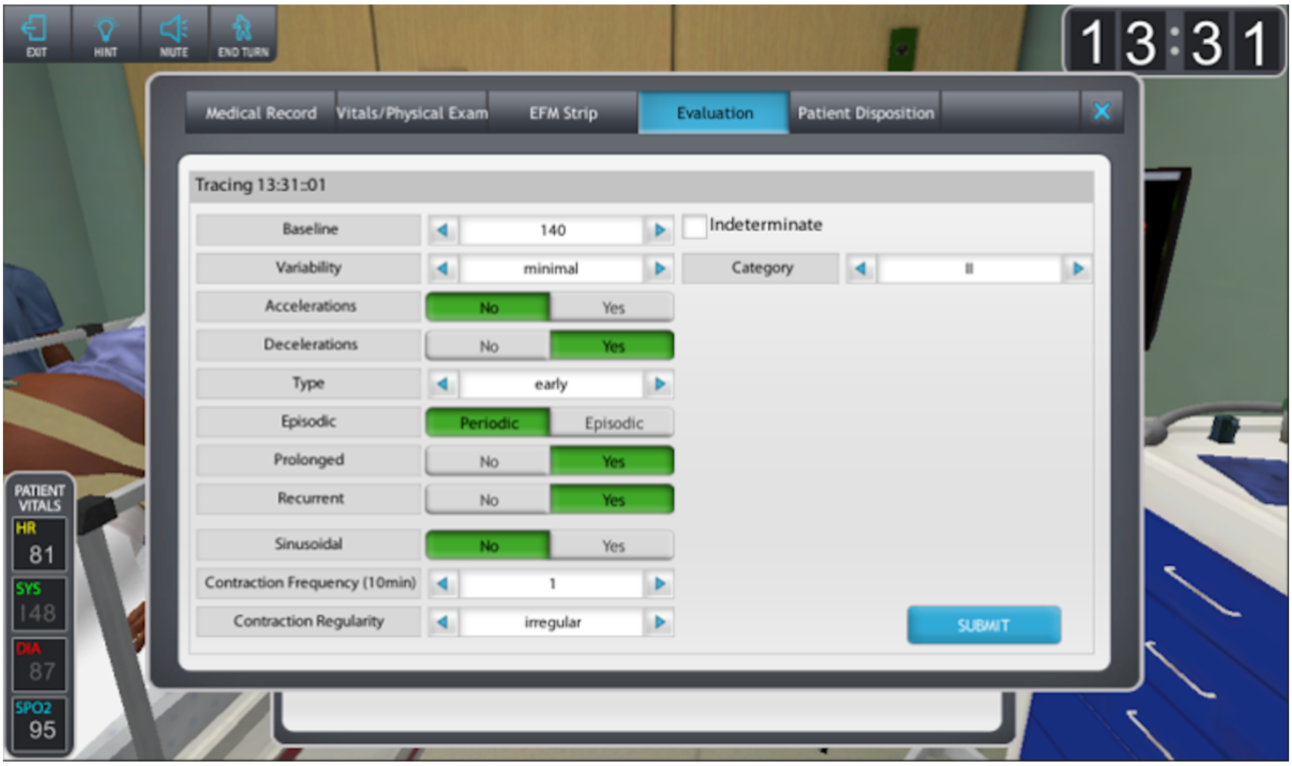
Participants were scored based on their verbalized interpretation of the EFM strip and description of subsequent actions. Experts in simulation and obstetrics (TA and SP) used a standardized scoring sheet to score the participants' interpretation and management. The scoring sheet was an objective tool that recorded what the participants verbalized; it has been used institutionally to create a mechanism to grade learners on their education. The observers recorded the category of tracing stated by the participant at each interpretation point. Then, the percent correct was calculated. At the end of each experience, the in-lab participants received a one-on-one debrief from the evaluator/facilitator on their interpretation of the EFM tracings and their overall decision making. The expert facilitators used the Plus/Delta debriefing model [26]. Participants were asked to identify what they felt they did well and what they would do differently if they participated in the scenario again. There was then a discussion around this, and the facilitator had an opportunity to add insight into items that they noticed were not mentioned. At the end of the debriefing sessions, participants were asked to identify several "take-aways" from the session. They were asked to think about how they will change their practice as a result of participating in this experiential activity.

Immediately following the training, participants completed a questionnaire using a computer provided in the simulation center. The questionnaire consisted of demographic information, questions regarding their experience with the simulation, and the opportunity to provide a free-text narrative of their experience.

Demographic information was collected, including employment position, years of experience, prior training experience, gender, and age. There were also Likert scale–based questions related to the participant's experience with different technologies, their perceptions of simulation-based training modalities, as well as their perceptions regarding the simulation training they had just completed. In addition, participants could provide free-text narratives regarding their experience.

Participants randomized to the Game Group were shown a video describing the functionality of the game before they could proceed to the training scenarios. The participant used their avatar to move from room to room to manage the patient. Participants utilized the game's controls to provide interpretations of the EFM strips and complete related patient care actions (example in Figure 4).

Figure 4. Depiction of the screen participants utilized to input answers in the serious game.



The game had an internal scoring mechanism equivalent to the in-lab scoring sheets. In the serious game, feedback was given based on interpretations of the fetal tracings as well as correct and incorrect actions completed by the learner. Incorrect actions triggered textual descriptions of feedback that pertained specifically to the incorrect action committed and mimicked the standardized feedback provided in the mannequin-based training.

Serious game participants completed a perceptions questionnaire identical to that completed by the mannequin-based training participants. Similarly, the questionnaire was completed on a screen to which they were directed immediately following their training.

Phase 2: Posttest Simulation Assessment

After participants completed Phase 1, there was a 3-6–month washout period before participants were assessed in Phase 2. To assess the EFM skills imparted to learners in both the Mannequin and Game Groups, participants were evaluated in an in-situ simulation scenario (described in previous sections) involving EFM interpretation and decision making. The scenarios for the in-situ simulation-based team training mirrored that of the mannequin-based training simulation. The scenarios contained different details from Phase 1 but were similar in the clinical topic and category of EFM tracings to evaluate as in the mannequin and serious game scenarios. Participants were first given instructions regarding the functionality of the mannequin and reminded to verbalize their EFM strip interpretations and care actions. Participants' verbalized answers were scored by a grader (the same graders used in the mannequin-based training scenarios) in real-time. The grader stood behind an opaque screen so they could not see the participant in an effort to blind the grader from the participant's allocation group. Correct answers to questions had been objectively determined in advance to reduce potential grader bias. Similar to Phase 1, subject matter experts and doctoral professionals with interest in assessments developed and scored the evaluations for the in-situ simulation of Phase 2.

Data Analysis

A Priori Power Analysis

Prior to beginning data collection, sample size calculations were conducted based on a noninferiority design (ie, Mannequin Group was not worse than the Game Group within a predetermined margin). The noninferiority margin was set to a –5-point difference in the primary outcome measure, namely performance on the MOST training. This analysis indicated that 64 participants per group (128 total) would be required to conclude noninferiority of the mannequin-based training.

Quantitative Data Analysis

Following numerous different recruitment efforts using various strategies, a total of 36 participants volunteered to be part of the study. While this sample size was not sufficient for testing noninferiority of the Mannequin Group, statistical analysis was performed to compare the performance of Mannequin Group participants to that of the Game Group participants on the posttest assessment simulation to detect if there were any significant differences that may be pertinent to understanding the results or to future studies. Statistical analysis was also utilized to compare differences in perceptions from the Likert scale questions in the posttraining assessment questionnaire. Descriptive statistics were performed to calculate mean, median, and SD of the performance and demographic data collected. Comparisons of continuous variables were conducted using student *t* tests or Wilcoxon rank sum tests for skewed data.

Categorical variables were compared using a chi square test or Fisher's exact test for small cell size (n<5). For statistical analysis, SAS 9.4 was utilized, and all analyses were carried out by a Masters-level biostatistician.

Qualitative Data Analysis

Qualitative analysis of the free-text narratives provided by participants during the posttraining questionnaire in Phase 1 was conducted utilizing thematic analysis [27]. Two researchers, one PhD-level, with extensive experience analyzing qualitative data (DH and NB) inductively elicited themes from the dataset. Inductive coding involves compiling themes as they emerge through the text, as opposed to deductive coding, which utilizes a previously developed set of themes [28]. A researcher with expertise in obstetrics (TA) then reviewed the themes and definitions for clinical accuracy and relevance. Once the themes had undergone subject matter expert review (TA), the two researchers (DH and NB) established interrater reliability with a Fleiss kappa value of 0.791, denoting substantial agreement [29]. The researchers then independently completed the coding. The completed coding was reviewed by both researchers to ensure consistency of coding. Finally, axial coding was utilized to collapse or combine similar themes [30].

Results

Characteristics of Study Participants

There were 36 total participants in the study. Table 1 provides demographic summary data for both participant groups. No significant differences were found in any of the demographic categories between the 2 treatment groups.

Table 1. Demographics of participants in the simulation (Mannequin Group) and serious game (Game Group) arms of the randomized controlled trial.

Characteristics	Game Group (n=18)	Mannequin Group (n=18)	*P* value
Participant role, n (%)			
Nurse	2 (11)	2 (11)	1.00
Resident physician	12 (67)	12 (67)	
Attending physician	4 (22)	4 (22)	
Gender (female), n (%)	16 (89)	14 (78)	.66
Age (years), mean (SD)	32.91 (7.35)	32.74 (7.4)	.41
Years in role, mean (SD)	2.83 (3.29)	2.32 (2.41)	.99

Quantitative Results

Posttest Performance Assessment

No significant differences were detected between the Mannequin and Game Groups in overall performance during Phase 2, namely the MOST scenario (measured as the percent correct; *P*=.43). Those in the Mannequin Group had an average score of 64.2%, while those in the Game Group had an average score of 53.8%.

Table 2 provides a breakdown of performance by EFM component. The Mannequin Group verbalized the correct contraction frequency more often (*P*=.04), and the Game Group was more likely to fail to state the contraction frequency (*P*=.04). None of the other tests demonstrated significant differences between groups.

Table 2. Performance in Phase 2 by electronic fetal monitoring (EFM) component, shown as the number and percent of participants providing the correct responses.

EFM component	Game Group (n=18), n (%)	Mannequin Group (n=18), n (%)	*P* value
Fetal heart rate category	14 (77.8)	9 (50.0)	.09
Fetal heart rate variability	6 (33.3)	9 (50.0)	.31
Declarations	15 (83.3)	16 (88.9)	1.00
Type	8 (44.4)	9 (50.0)	.74
Baseline fetal heart rate	8 (44.4)	8 (44.4)	1.00
Accelerations	11 (61.1)	11 (61.1)	1.00
Contraction frequency	5 (27.8)	11 (61.1)	.04
Recurrent	7 (40.0)	18 (100)	.06
Regularity	13 (72.2)	13 (72.2)	1.00

Post-Training Perceptions Questionnaire

Table 3 provides a comparison of the 2 participant groups for the 5 survey questions that pertained specifically to the participants' experience during their training intervention. There were no significant differences in perceptions between groups.

Table 3. Simulation training perception survey results, reported as the average Likert scale responses on scale of 1 (strongly disagree) to 5 (strongly agree).

Survey prompt	Game Group, mean	Mannequin Group, mean	*P* value
Learning through simulation as I just experienced is frustrating.	3.11	2.33	.19
I think the simulation training I just experienced is a great tool for learning.	4.11	4.50	.46
I would like to be involved in more simulation training of the sort I just experienced.	4.00	4.11	.16
I do not see how the kind of training I just experienced is relevant to my work.	2.00	1.47	.37
I am satisfied with this training experience.	3.83	4.06	.82

Qualitative Results

Summary

Of the 36 total participants, 16 (44%) voluntarily provided free-text narratives during the survey portion of Phase 1 of the study to describe their experience. From the participants' descriptions, 6 unique themes emerged. We describe each theme in the following sections, including participant quotes, followed by a participant ID displayed such that the first letter signals the participant role (A: attending physician, R: resident physician, N: nurse), followed by an ID number, with the last letter denoting the treatment group (M: Mannequin Group, G: Game Group). Themes are listed in descending order from most to least commonly discussed by participants.

Overall Experience and Opinions on Simulation

Participants from both the Mannequin and Game Groups provided feedback regarding their overall experience and opinions related to simulation, which were positive, neutral, or negative. Game Group participants, for example, noted that they would "rather do this then listen to a lecture" [R-10-G]. Another described that "this would be a useful tool once you get used to what is expected" [A-07-G].

Mannequin Group participants also provided feedback about the simulation: "I liked the concept of following a patient through her labor course" [A-09-M]. Mannequin Group participants also tended to describe their views of the simulation, which varied by participant. One stated: "I strongly believe that simulation-based training is absolutely vital to training for any and all medical procedures" [N-06-M]. Conversely, another participant described, "I find simulations frustrating for their lack of natural conclusion and the feeling of being judged" [R-23-M].

Realism of the Simulation

Both Mannequin and Game Group participants described shortcomings related to the realism of the simulation. One Game Group participant said "I felt like I have to keep doing things and waiting for something to happen or change though that is not what I would have done in real life" [R-05-G]. Mannequin Group participants also expressed concerns, such as "It was a bit random to be having to act like I am interacting with a patient but then randomly having to do a strip review when talking to the patient" [A-09-M].

Participants also described elements that could be adapted to improve realism, such as:

> *Picking a [EFM] strip most like our home institution in terms of size and markings.* [R-07-G]
>
> *I never push IV meds, so I am not sure why I can't just call out a medication.* [A-09-M]

Navigating the Simulation and Technology Issues

Participants, predominantly from the Game Group, highlighted issues and areas for improvement related to navigating the simulation, including:

> *I would have liked to have a practice scenario to figure out how to virtually assess my patient.* [N-03-G]
>
> *There is no back button to reverse steps made (if done in error).* [R-09-G]

Another Game Group participant also reported that "some of the actions I performed were not recorded in the summary, and I did not get credit for them" [R-14-G].

Simulation as a Learning Tool

Multiple participants discussed the value of simulation as a learning tool, although this was more commonly discussed by participants in the Mannequin Group. Feedback included:

> *[I] look forward to improving my virtual assessment skills and score.* [N-03-G]
>
> *Having gone through the motions in sim lab, when the stakes weren't as high, provides comfort and give you confidence when faced with the real scenario.* [R-22-M]

Facilitators and Challenges to Understanding Simulation Feedback

Related to learning, one Game Group participant described challenges in understanding the feedback they were given via a "score card":

> *The review score card did not help me understand where my clinical decision making went wrong.* [N-03-G]

Alternatively, one of the Mannequin Group participants noted that "in real life, it is helpful to have on-the-spot feedback" [R-02-M].

Mismatch in Role and Simulation Activities

Lastly, 2 nurses, 1 from each treatment group, described that they were asked to do things that fall outside of their scope of practice:

> *Decision to call a c/s [caesarean section] is not in my scope of practice as an L&D [labor and delivery] RN.* [N-03-G]
>
> *Post-partum nurses rarely have to decide what to do with MgSo4 toxicity.* [N-05-M]

Discussion

This study provides lessons to help advance the future of serious game simulation techniques through qualitative and quantitative evidence. Among participants in both groups, the difference in final correct answers was not shown to be statistically significant. The serious game training tool developed has potential utility in providing education to those without access to large simulation centers; however, further validation is needed to demonstrate if this tool is as effective as mannequin-based simulation. Additional lessons were provided through the assessment of participant perceptions. Namely, feedback suggested participants were accepting of and satisfied with the new training modality. Further, open-ended feedback elicited important themes for improvement to advance the future of serious games as a training tool.

For individual components of EFM interpretation and decision making, the only component that showed statistical significance was interpretation of the uterine contraction pattern. Those in the Mannequin Group performed better. However, those in the Game Group had a significantly higher occurrence of "no answer" (ie, they failed to verbalize an answer for this component), which was counted as an incorrect response. We hypothesize that the Mannequin Group gained experience with verbalizing the contraction pattern during their training and were therefore more comfortable with this during Phase 2. Those in the Game Group had no practice with verbalization until the posttest assessment. This finding highlights a type of experience gained from a simulation center that may be difficult to replicate in serious games and needs to be considered in the design of future studies.

Subjective responses, including those from the questionnaires and narrative feedback (see Qualitative Analysis), highlighted positives and room for improvement in both groups. Notably, in the perceptions survey, both groups stated the simulation modality they had experienced was a great tool for learning, they wanted to be involved in that simulation modality more, and they were satisfied with their training. This suggests that participants were generally accepting of the new training modality (serious game–based simulation). These findings were also reflected in the qualitative analysis themes related to "overall experience" and "simulation as learning tool."

Qualitative feedback from the serious game group showed that some users found the game challenging to navigate and experienced technical difficulties. Suggestions to improve serious game play include providing practice scenarios, providing better affordances related to how to complete a scenario, and incorporating an "undo" feature. The Game Group did watch a mandatory video, but the finding related to practice sessions suggests interactive tutorials and some type of competency assessment prior to using the training game may be beneficial. In general, maneuvering through the game must be simple. We completed a usability study prior to the assessment reported here, but these new findings underscore the importance of user-centered design as an iterative process that occurs throughout the product development lifecycle. Designing serious gaming experiences to suit user needs and goals may encourage self-guided mastery of competencies, particularly in an area as crucial and challenging as EFM interpretation.

Participants from both groups described challenges related to realism of the simulation. One participant suggested that the serious game–based simulation may benefit from using EFM strips similar to what is used in their hospital. This highlights an important point for expanding serious game use to national and international platforms. Specifically, it suggests that it may be beneficial to offer customization settings so the game feels as realistic as possible. Nurse participants also described that some of the actions fell outside of their scope of practice. Relatively few studies have incorporated nurses into serious game development and assessment [16-19], and it is important to ensure participants of various roles have opportunities for serious game–based learning. However, our finding highlights the need for adapted scenarios that accurately match the responsibilities of different roles.

The novelty of serious games for delivery of health care education will have to endure a learning curve in learners' comfort levels of navigation. This learning curve can be mitigated through iterative testing to ensure trainings are usable for the various end-user groups. Over time, serious games for health care delivery have the potential to become mainstream, much like mannequin-based simulation has over the past 10-15 years. In traditional simulation education, it is important that the level of fidelity or realism of the training matches the objectives of the training with regards to the patient, clinical facilities, and clinical scenario [31]. Understanding levels of realism across the patient, clinical facilities, and clinical scenarios necessary to meet objectives related in the serious game format will need to be similarly considered as this type of simulation training gains traction in the future.

Limitations

Our study was limited by the difficulty in recruitment of participants. No statistically significant differences were detected in the primary outcome measure, but sufficient power was not attained to conclude equivalence of the original training method (mannequin-based simulation) to the serious game–based training. Further investigation is needed to determine if the methods are equivalent, but this study certainly supports the continued investigation of the value of virtual simulation environments such as serious game–based scenarios as alternatives to traditional mannequin-based simulation. While all participants had a washout period of at least 3 months, due to their schedules, some had washout periods as long as 6 months. This had an unknown effect on outcomes. Future studies may track this period as a covariate to test its potential effect on outcome measures.

The perception-based questionnaires were pilot tested but were not psychometrically validated, which should be considered in the interpretation of these results.

Another potential limitation may be the possible training effect from both methodologies. Though not a desired outcome, we recognize that there is sustained difficulty in interpretation of EFM regardless of type of training. Though years of experience may help, there is still significant difficulty in predicting fetal wellbeing in all EFM situations, particularly category 2 tracings [32,33].

The qualitative analysis was limited in that free-text narratives were not mandatory and only provided by 44% of the participants. Future studies may employ more robust techniques, such as semistructured interviews, to gain more comprehensive feedback from all participants. The discovered qualitative themes, however, still provide important insight into improving the future of serious game–based simulation.

This study was conducted specifically with OBGYN providers and nurses, which may not necessarily be generalizable to other clinical subspecialties.

Conclusions

Data indicated that the serious game is viewed as effective both by physician and nurse participants, and quantitative measures suggest that serious game participants will have similar performance to those participating in the human-patient simulation. However, the study was not sufficiently powered to assess our hypotheses. Larger studies are necessary before definitive conclusions can be made about how serious game training compares to the more well-established mannequin-based training methods in obstetrics as well as other specialties. We have demonstrated feasibility of using serious game training to deliver education as an alternative to simulation-based training and provided insight to improve these methods in future implementations of this technology. Our study has indicated that a focus on realism and usability of the training tool will be important areas of focus in serious game development in the future.

Acknowledgments

The authors would like to thank Sara R Parker, MD, Mihriye Mete, PhD, and Sameer Desale, MS for their contributions to data collection and analysis. This project was supported by grant number R18HS021444 from the Agency for Healthcare Research and Quality.

Conflicts of Interest

None declared.

This randomized study was not registered. The authors explained that their study "involved health professional participants (not patients or lay participants) and the educational outcome assessed was related to the health professionals. The study was funded by the Agency for Healthcare Research and Quality who did not require it be registered as a clinical trial." The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because only health professionals were involved in the study. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

References

1. American College of Obstetricians and Gynecologists Committee on Patient SafetyQuality Improvement. Committee opinion no. 590: preparing for clinical emergencies in obstetrics and gynecology. Obstet Gynecol 2014 Mar;123(3):722-725. [doi: 10.1097/01.AOG.0000444442.04111.c6] [Medline: 24553170]
2. American College of Obstetricians and Gynecologists Committee on Patient Safety and Quality Improvement. ACOG Committee Opinion No. 447: Patient Safety in Obstetrics and Gynecology. 2009;114(6):1424-1427. [doi: 10.1097/aog.0b013e3181c6f90e]
3. Weinstock PH, Kappus LJ, Kleinman ME, Grenier B, Hickey P, Burns JP. Toward a new paradigm in hospital-based pediatric education: the development of an onsite simulator program. Pediatr Crit Care Med 2005 Nov;6(6):635-641. [doi: 10.1097/01.pcc.0000185489.07469.af] [Medline: 16276327]
4. Weinstock PH, Kappus LJ, Garden A, Burns JP. Simulation at the point of care: Reduced-cost, in situ training via a mobile cart. Pediatric Critical Care Medicine 2009;10(2):176-181. [doi: 10.1097/pcc.0b013e3181956c6f]
5. Rodgers D. High-fidelity patient simulation: a descriptive white paper report. Healthcare Simulation Strategies. 2007. URL: https://sites.psu.edu/hersheysimulation/files/2016/11/Simulation-White-Paper-Rodgers-1pk08yn.pdf [accessed 2020-09-04]
6. Morgan PJ, Cleave-Hogg D. A worldwide survey of the use of simulation in anesthesia. Can J Anesth/J Can Anesth 2002 Aug;49(7):659-662. [doi: 10.1007/bf03017441]
7. Maheu-Cadotte M, Cossette S, Dubé V, Fontaine G, Mailhot T, Lavoie P, et al. Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: a systematic review and meta-analysis protocol. BMJ Open 2018 Mar 16;8(3):e019871. [doi: 10.1136/bmjopen-2017-019871] [Medline: 29549206]

8. Ricciardi F, De Paolis LT. A Comprehensive Review of Serious Games in Health Professions. International Journal of Computer Games Technology 2014;2014:1-11. [doi: 10.1155/2014/787968]
9. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a Virtual Patient Trainer Teach Student Nurses How to Save Lives—Teaching Nursing Students About Pediatric Respiratory Diseases. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare 2012;7(1):10-17. [doi: 10.1097/sih.0b013e31823652de]
10. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
11. Dankbaar MEW, Richters O, Kalkman CJ, Prins G, Ten Cate OTJ, van Merrienboer JJG, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. BMC Med Educ 2017 Feb 02;17(1):30 [FREE Full text] [doi: 10.1186/s12909-016-0836-5] [Medline: 28148296]
12. Knight JF, Carley S, Tregunna B, Jarvis S, Smithies R, de Freitas S, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. Resuscitation 2010 Sep;81(9):1175-1179. [doi: 10.1016/j.resuscitation.2010.03.042] [Medline: 20732609]
13. Graafland M, Bemelman WA, Schijven MP. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. Surg Endosc 2017 Oct 9;31(10):4093-4101 [FREE Full text] [doi: 10.1007/s00464-017-5456-6] [Medline: 28281117]
14. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho ICM. InsuOnline, an Electronic Game for Medical Education on Insulin Therapy: A Randomized Controlled Trial With Primary Care Physicians. J Med Internet Res 2017 Mar 09;19(3):e72 [FREE Full text] [doi: 10.2196/jmir.6944] [Medline: 28279950]
15. Cendan JC, Johnson TR. Enhancing learning through optimal sequencing of web-based and manikin simulators to teach shock physiology in the medical curriculum. Adv Physiol Educ 2011 Dec;35(4):402-407. [doi: 10.1152/advan.00061.2011] [Medline: 22139778]
16. Ijaz A, Khan MY, Ali SM, Qadir J, Boulos MNK. Serious games for healthcare professional training: A systematic review. European Journal of Biomedical Informatics 2019;15(1):1 [FREE Full text]
17. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health Education Serious Games Targeting Health Care Providers, Patients, and Public Health Users: Scoping Review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
18. Chon S, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious Games in Surgical Medical Education: A Virtual Emergency Department as a Tool for Teaching Clinical Reasoning to Medical Students. JMIR Serious Games 2019 Mar 05;7(1):e13028 [FREE Full text] [doi: 10.2196/13028] [Medline: 30835239]
19. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec 19;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
20. Jena AB, Seabury S, Lakdawalla D, Chandra A. Malpractice Risk According to Physician Specialty. N Engl J Med 2011 Aug 18;365(7):629-636. [doi: 10.1056/nejmsa1012370]
21. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. Med Educ 2006 Aug;40(8):792-797. [doi: 10.1111/j.1365-2929.2006.02528.x] [Medline: 16869926]
22. Deering S, Johnston LC, Colacchio K. Multidisciplinary teamwork and communication training. Semin Perinatol 2011 Apr;35(2):89-96. [doi: 10.1053/j.semperi.2011.01.009] [Medline: 21440817]
23. Merién AR, van de Ven J, Mol B, Houterman S, Oei S. Multidisciplinary Team Training in a Simulation Setting for Acute Obstetric Emergencies. 2010;115(5):1021-1031. [doi: 10.1097/aog.0b013e3181d9f4cd]
24. Wayne DB, Butter J, Siddall VJ, Fudala MJ, Linquist LA, Feinglass J, et al. Simulation-based training of internal medicine residents in advanced cardiac life support protocols: a randomized trial. Teach Learn Med 2005 Jul;17(3):210-216. [doi: 10.1207/s15328015tlm1703_3] [Medline: 16042514]
25. Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. Chest 2008 Jan;133(1):56-61. [doi: 10.1378/chest.07-0131] [Medline: 17573509]
26. Brown M, Holt R. Utilizing Plus/Delta Debriefing to Enhance Learning in Phlebotomy Simulations. Am J Clin Pathol 2015 Oct 01;144(suppl 2):A107-A107. [doi: 10.1093/ajcp/144.suppl2.107]
27. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Newbury Park, CA: Sage Publications, Inc; 2015.
28. Miles M, Huberman A, Saldana J. Qualitative Data Analysis: A Methods Sourcebook. Newbury Park, CA: Sage Publications, Inc; 2014.
29. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
30. Saldana J. The Coding Manual for Qualitative Researchers. 2nd ed. Newbury Park, CA: Sage Publications, Inc; 2015.

31. Tun JK, Alinier G, Tang J, Kneebone RL. Redefining Simulation Fidelity for Healthcare Education. 2015 Apr 05;46(2):159-174. [doi: 10.1177/1046878115576103]
32. Shepherd E, Salam R, Middleton P, Makrides M, McIntyre S, Badawi N, et al. Antenatal and intrapartum interventions for preventing cerebral palsy: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2017 Aug 08;8:CD012077 [FREE Full text] [doi: 10.1002/14651858.CD012077.pub2] [Medline: 28786098]
33. Macones G, Hankins G, Spong C, Hauth J, Moore T. The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. J Obstet Gynecol Neonatal Nurs 2008;37(5):510-515. [doi: 10.1111/j.1552-6909.2008.00284.x] [Medline: 18761565]

Abbreviations

EFM: electronic fetal monitoring
MOST: MedStar Obstetrical Safety Training
OBGYN: obstetrician gynecologist

Edited by G Eysenbach; submitted 07.06.20; peer-reviewed by M Tara, K Glover; comments to author 24.07.20; revised version received 28.07.20; accepted 03.08.20; published 28.09.20.

Please cite as:
Benda NC, Kellogg KM, Hoffman DJ, Fairbanks RJ, Auguste T
Lessons Learned From an Evaluation of Serious Gaming as an Alternative to Mannequin-Based Simulation Technology: Randomized Controlled Trial
JMIR Serious Games 2020;8(3):e21123
URL: http://games.jmir.org/2020/3/e21123/
doi:10.2196/21123
PMID:32985993

Original Paper

Assessing Ethoshunt as a Gamification-Based Mobile App in Ethics Education: Pilot Mixed-Methods Study

Noor Syamilah Zakaria[1], BEd, MEd, PhD; M Iqbal Saripan[2], BEng, PhD; Neerushah Subarimaniam[1], BSocSc, MSc; Alyani Ismail[2], BEng, MSc, PhD

[1]Department of Counselor Education and Counseling Psychology, Faculty of Educational Studies, Universiti Putra Malaysia, Selangor Darul Ehsan, Malaysia

[2]Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Selangor Darul Ehsan, Malaysia

Corresponding Author:
Noor Syamilah Zakaria, BEd, MEd, PhD
Department of Counselor Education and Counseling Psychology
Faculty of Educational Studies
Universiti Putra Malaysia
UPM Serdang
Selangor Darul Ehsan, 43400
Malaysia
Phone: 60 127849776
Email: syamilah@upm.edu.my

Abstract

Background: Gamification has remarkable potential in the learning space. The process of creating a gamified system and its influence on human behavior reflect the interaction between educators and machines.

Objective: The purpose of this pilot study was to present Ethoshunt as a gamification-based mobile app that can be used in teaching and learning ethics.

Methods: This study involved a mixed-methods research design. The researchers surveyed 39 undergraduate students who were introduced to Ethoshunt in order to examine the relationships between mobile app usability and positive emotions, ethical competency, and user experience. Affinity diagramming was used as a tool to organize the opinions and experiences of participants using featured gamification elements.

Results: Game dynamics and game mechanics explained the functionality of Ethoshunt. In addition, the learning flow through Ethoshunt was discussed. Overall, the findings were positive, and mobile app usability had the strongest relationship with positive emotions (r=0.744, P<.001), followed by ethical competency (r=0.686, P<.001) and user experience (r=0.614, P<.001).

Conclusions: Positive emotions could be perceived as an important dimension in the development and usability of Ethoshunt. The researchers suggest that the gamification-based mobile app advocated in this study may provide ideas for ethics educators who wish to develop a technology-mediated learning environment.

(JMIR Serious Games 2020;8(3):e18247) doi:10.2196/18247

KEYWORDS

gamification; ethics; education; ethics education; ethical competency; mobile app; mobile app usability

Introduction

Background

Gamification is the use of a game-like concept to create engaging and supportive learning environments that motivate students. Elements of games have been historically important in teaching and learning. Game elements effectively engage students, and the primary role of educators to engage students in learning is becoming a priority [1-3]. According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), the process of integrating new tools and technologies that involve game elements depends on the educator's ability to: (1) construct a learning environment by employing new ideas; (2) consolidate new technology with a new pedagogy; and (3) promote socially active classrooms to encourage group work, cooperative interaction, and collaborative learning [4]. Furthermore, integrating game elements in education is a technique that is used to foster collaboration among students [5] and to extend learning while developing their cognitive, emotional, and social dimensions [6].

The application of game elements, such as game dynamics and game mechanics, commonly known as "gamification" has been entering the realms of education, lifestyle, and health [7]. In the education context, gamification is regarded as a tool that has been influential [8] and a prominent trend in recent years [9]. It creates a great deal of value for students since it encourages learning through entertainment [10]. With multifunction tools, gamification is defined as the use of game design elements in nongame contexts [11]. The entire process of gamification is to create interactions between educators and machines, but not to replace educators [11].

Gamification is an application that integrates both game mechanics and game dynamics with the support of a virtual learning environment, offering various benefits to humans [12]. Previous studies have investigated the development and usage of gamification across disciplines. For instance, gamification positively influences consumers' behaviors in switching intention between membership cards and mobile apps [13]. On the other hand, gamification increases bicycle riding activities and logging behaviors through a point system [14], promotes the quality of teacher education and in-service training, assists researchers in environmental multicriteria decision analysis specifically related to water [15], and supports collaborative learning in primary education through implementation of a tactile and tangible multitablet gamified quiz system [2].

Gamification researchers have also explored personality types that can predict players' preferences for game elements and mechanics [7], as well as energy-related behaviors that are necessary to engage residential customers in energy application enabled by smart meters [16]. Interestingly, many researchers have designed gamification-based applications to capture consumers' fun and "gameful" experiences [9] and to observe weight loss in overweight and obese adolescents, which eventually improves their health conditions [12]. On the other hand, Zhang et al [17] used gamification for cognitive bias modification. Besides usage, usability of gamification is equally important as it enables users to perform tasks effectively and efficiently. Usability ensures a gamification-based tool impacts users in many ways, such as achieving goals, increasing proactive behaviors, and improving performance; otherwise, the tool would be worthless.

Similarly, usability was the main concern in this study. It is critically important to measure the usability of a gamification-based tool to ensure it is usable, useful, and effective in teaching and learning ethics. Therefore, the aim of this study was to present the development of a gamification-based mobile app called Ethoshunt and assessment of its usability in ethics education. The researchers used a correlation method to examine the relationships between mobile app usability and positive emotions, ethical competency, and user experience. Opinions from students based on their experiences of using Ethoshunt were also gathered. The results were expected to provide evidence of the effectiveness and usability of Ethoshunt.

Related Work

Gamification in Ethics Education

Gamification has a large potential in teaching and learning activities. Furthermore, gamification strengthens teaching and learning processes [18]. Through gamification, students as gamification-based application users potentially exhibit positive learning behaviors. In addition, gamification is almost everywhere in the form of puzzles; adventure, simulation, edutainment, and strategy games; and real-time strategy games [19]. Gamification reflects a shifting idea of using game design elements, such as the mechanics and dynamics of games, in nongame themes, products, contexts, and services to make difficult tasks easier [10].

Gamification is an effective engagement tool that reinforces human behaviors through specific educational activities with the assistance of game mechanics, such as levels, ranks, missions, and instant rewards [20]. The most used game mechanics are badges, points, levels, prizes, storytelling, progress bars, and leaderboards [21-23]. For instance, an ethics educator may not conduct formative assessments in tutorial classes, and instead, may replace it with gamification-based assessments inclusive of prizes that would create game-based experiences in educational contexts. These experiences would increase students' relatedness, competence, and autonomy [19,24]. Furthermore, gamification is pleasing, engaging [17,25], and sustainable, as it provides an interactive educational environment. Education-based gamification systems, such as GamiCad, Jigsaw, and Gamified Multimedia Content Production, guarantee students' quick task completion during the process of reviewing the lessons learned [26].

Gamification allows students to experience a feeling of energized focus, enjoy the process of classroom activity, and get involved in the learning process [2]. In addition, gamification can be utilized as a motivational tool [1,8,27], promotes a flexible learning environment, and provides challenges to high-ability students [12]. In addition, students can maintain positive attitudes toward courses integrated with gamification, and gamification can influence students' learning and achievements [28]. Students are generally enthusiastic about gamification and learn more with gamification. The previous findings reflect that nontraditional teaching methods integrating technology-mediated learning can enhance students' engagement in the classroom. However, it is important to recruit quality educators, as they are expected to make efficient use of technical infrastructure, use technology effectively, enable effective communication, adapt gamification content to the ethics course, and interpret the outcome of gamification-based applications [29].

An alternative concept to build a motivating nature in game-based pedagogy is ethics educators' competencies. Besides effectively making use of the technology, ethics educators are expected to be competent in four areas, including pedagogical, collaborative, technological, and creative areas [20]. A pedagogically competent educator ought to have the ability to engage students in planning and designing gamification, requiring collaboration (second area) within the school and other schools. Moreover, educators must be technologically competent, as they will need to analyze games and technological tools. It is expected as well that technologically competent educators will be able to overcome technology-related obstacles [20]. Finally, creative educators are needed to create playful learning zones to explore and enhance gamification in education. The creativity aspect of teaching would assist in sustaining students' motivation and promote self-development. For the motivation aspect, gamification promotes behavioral changes. Timpel et al [12] mentioned that motivation is the heart of gamification. Gamification works best to boost motivation through engagement, which can make the most challenging and tedious tasks enjoyable [27]. Timpel et al [12] further revealed that the aspect of motivation in gamification can be explained further according to the self-determination theory and organismic integration theory.

The self-determination theory focuses on what drives students to make choices without external influences and argues that students or gamification users are self-motivated and self-determined as the gamified system provides them a sense of social relatedness and autonomy [12]. The elements in gamification ultimately create a state of intrinsic motivation that influences users to go through challenges and make achievements. According to Su and Cheng [22], intrinsic motivation that is developed based on feelings of competence and autonomy will keep students engaged and help them learn actively for many hours with no rewards. On the other hand, the organismic integration theory describes different types of extrinsic motivation and highlights that behavioral regulations are experienced as comparatively alien to the self [30]. In other words, individuals' actions are compelled by externally controlled rewards or punishments. For example, students who are exposed to gamification in conservative classroom lectures will feel motivated, specifically when they are rewarded with progress bars, points, or badges, and the motivation indirectly builds collaboration with other learning peers.

Gamification has the potential to transform an extrinsic experience into internal meaning and help students to reflect upon and reorganize their familiarities [31,32]. However, in the long run, the reward mechanism, which also serves as a controlling aspect, may erode intrinsic motivation [12]. Thus, it is essential to select meaningful educational game elements that are in line with the users' interests and goals to sustain their intrinsic motivation. According to Rodrigues et al [2], points, quizzes, and challenges are the most used game elements.

Assessing Positive Emotions, Ethical Competency, User Experience, and Mobile App Usability

Positive emotions refer to positive effects functioning as internal signals that increase the motivation level [33]. In addition, positive emotions, such as joy, contentment, love, and interest, are markers of optimal well-being or flourishing. Positive emotions are recognizable via vocal, visual, auditory, touch, and facial cues [34]. In the context of this study, gamification has a tendency to provide positive emotions that may break existing habits and substitute the habitual behaviors with new stabilized behaviors [35]. Gamification invokes students' behaviors in the classroom by activating their motivation [36] and sustaining focus in the classroom. For example, students who are exposed to gamification in the classroom can learn in a fun environment while sharing their experiences with educators and learning peers. The game dynamics and game mechanics in gamification-based tools attract students, as they increase students' positive emotions, which rationally makes them invest more time in exploring the game, boosts motivation, and fosters their curiosity to know more about the game. According to Korn et al [36], gamification is able to shift the students' mood spectrum toward a more relaxed condition. The delightful concept of gamification involves more emotions; however, the happy and unhappy emotions balance each other [36].

Ethical competency refers to ethical behaviors and actions that require ethical knowledge and reflection [37]. Ethical competency aids in achieving organizational goals, improves individuals' performances, improves quality of the services provided, and invites positive consequences to the organization [38]. Ethical competency is critical to the well-being of students. Living ethically can be articulated as projecting the highest standard of the belief system. To comprehensively teach ethics education in a semester is a challenge; nonetheless, a foundation of knowledge, thinking, and acting can be set in place [26,32,39,40]. In this study, the ethics education instructor used gamification for disseminating knowledge from the literature and text books, dialoguing in conservative classroom lectures regarding knowledge application, and utilizing experiential activities to apply knowledge in real-life situations.

On the other hand, user experience is defined as an episode that involves interactions between categories as follows: (1) an individual and an individual; (2) an individual and a service provider; and (3) a customer and a product provider resulting in a response [8]. Sheng and Teo [41] mentioned that user experience reflects an individual's expectations and the stimuli that result from interactions between the categories noted previously [8]. The researchers also mentioned that experiences perceived by the students will be a complex feeling, and usually, it is difficult to distinguish between two different experiences. In this study, user experience refers to what is perceived by the students after being exposed to gamification in conservative classroom lectures of ethics education. Experience can be described in the form of feelings, emotions, behaviors, or mental state.

Finally, mobile app usability refers to a method of measuring a product's ease of use [42]. Usability reflects the focus for every phase of design to ensure a mobile app is designed well, with usability in mind throughout the designing process [43]. For example, a mobile app that is designed well and fits the current needs of people would achieve high usability. However, app type and the location context may affect users' mobile experiences [44], which eventually would contribute to poor usability. This study aimed to measure the usability of Ethoshunt as one of the constructs to ensure the app promotes fun and a meaningful learning experience of ethics education. Students are encouraged to explore Ethoshunt as a user-friendly tool that increases their chance of performing well in an ethics education course.

There are limited definitions and citations specifically related to correlations among positive emotions, ethical competency, user experience, and mobile app usability. This condition reflects the ways they can be integrated in gamification-related studies. The researchers found that the four constructs were established in gamification-related literature by many researchers, and it is critically important to measure mobile app usability to determine the usability of Ethoshunt. Thus, the four constructs fit the purpose of this study, which is primarily to present Ethoshunt as a gamification tool in ethics education. In this study, the researchers aimed to develop a mobile app involving game elements integrating the dynamics and mechanics of gamification. The mobile app Ethoshunt can store ethos points and ethos levels through hints sent by the ethics educator. Furthermore, this paper highlights the relationships between mobile app usability and positive emotions, ethical competency, and user experience, which were measured on a scale based on the developed mobile app. Finally, the researchers analyzed five open-ended questions to further explore participants' personal experiences in using Ethoshunt.

Methods

Ethoshunt Development Process

The development of Ethoshunt is divided into four steps. The first step is the identification of the game mechanics and game dynamics of Ethoshunt. The second step is the definition of the components of the overall mobile app process module workflow. The third step is the presentation of how Ethoshunt functions. Finally, the fourth step is the identification of the learning flow through Ethoshunt integration.

Participants

The participants were students enrolled in a counseling ethics education course at a public university in Malaysia. The students were all first semester undergraduate students who were diverse in terms of age, gender, race, and cultural identity. The participants were chosen based on the research purpose and consideration of the comprehensiveness of the resources available. Ethics education is a cut and dry course. The course is considered complex and covers a large number of topics. The students registered for the counseling ethics education course are required to read, digest, and memorize the contents included in the Malaysian Counselor Act 1998 (Act 580) and Counselors Code of Ethics [45]. They are also required to attend lectures that are conducted traditionally in a classroom setting and complete assignments given by their educator. Therefore, a change is needed to ensure that student motivation to learn ethics is sustained and to provide an engaging learning environment.

A total of 39 undergraduate counseling students participated in this pilot study. They were selected through cluster sampling, in which a particular batch of students was chosen as participants. The Ethoshunt development project was conducted throughout four full semesters, which is equivalent to 2 years. On the other hand, the ethics education course is offered once in a year, resulting in two batches or clusters of students for 2 continuous years. Each cluster consists of students studying in the following two programs: (1) Bachelor of Education in Guidance and Counseling with Honors and (2) Bachelor of Counseling with Honors. Therefore, one cluster of students was selected randomly as participants of this study, resulting in 39 counseling students. The population consisted of 84 participants.

Data Analysis

Besides development of Ethoshunt, the researchers employed a mixed-methods research design to collect and analyze both quantitative and qualitative data gathered from a sample introduced to Ethoshunt. The researchers conducted a pilot study to assess the extent of positive emotions, ethical competency, user experience, and mobile app usability among undergraduate counseling students introduced to Ethoshunt and to capture their experiences of using Ethoshunt. The first stage of the pilot study involved survey analysis using Pearson correlation, and the purpose was to examine the relationships between two constructs as follows: (1) positive emotions and mobile app usability; (2) ethical competency and mobile app usability; and (3) user experience and mobile app usability. This was also called the quantitative phase.

The Gamification User Scale, a Malay language Likert-type scale ranging from 5 (strongly agree) to 1 (strongly disagree), was developed to measure positive emotions, ethical competency, user experience, and mobile app usability. Positive emotions, ethical competency, user experience, and mobile app usability are grouped into four sections, and the scale consists of 20 items. The first section consists of four items, and it measures students' positive emotions regarding Ethoshunt. The second section consists of four items about mobile app usability. The third section consists of eight items on students' ethical competency, and the final section involves four items that measure user experience. The scale has a high internal consistency (reliability, Cronbach α coefficient of .94). A panel of experts evaluated the content validity of the scale. Three experts, who are registered counselors with the Board of Counselors (Malaysia), evaluated the content validity of the scale in terms of comprehensiveness and clarity of the items. The examples of items for each construct are as follows:

- Positive emotions: (1) Ethoshunt is fun and (2) I am excited to learn while playing.
- Ethical competency: (1) Ethoshunt enhances my understanding in learning counseling ethics and (2) Ethoshunt assists me in application of counseling ethics education.
- User experience: (1) I am interested to learn using a mobile app in the future and (2) Ethoshunt makes me think about personal well-being.
- Mobile app usability: (1) A mobile app can make learning of counseling ethics more fun and (2) A mobile app can replace the traditional method of learning counseling ethics.

Another separate sheet was attached to the scale, and it involved the second stage of analysis, which is the qualitative phase. The separate sheet sought to collect students' opinions based on their experiences using featured gamification elements. The researchers conducted analyses manually to understand students' insights. An affinity diagram was used as a tool to analyze qualitative data by integrating the traditional qualitative simple data analysis method of using a pen and sticky notes. Opinions gathered from the students were transformed into an affinity diagram involving the following three steps: (1) capture; (2) group; and (3) label. The opinions from all participants were captured, grouped, and labeled to determine students' insights regarding Ethoshunt as a gamification tool in ethics education. A total of five open-ended questions were included in the questionnaire, in addition to the scale (Multimedia Appendix 1). All opinions provided by the participants were coded using a traditional simple qualitative data analysis tool.

Data Collection

The developed Ethoshunt app was used in parallel with conservative classroom lectures. It was not meant to replace the lectures, but rather complement them in a fun way. The ethics course instructor conducted the conservative classroom lectures for the first three topics. Thereafter, the instructor started the first set of hints, and the students followed the hints to get their ethos points, which eventually brought them to the next level of hints. The Ethoshunt was used until week 12 of the semester academic calendar. The ethics course instructor then rewarded the students who successfully managed to get to the next level up to the matured adult level, with treats, educational visits, or educational cash vouchers. Based on the developed Ethoshunt, the students' positive emotions, ethical competency, user experience, and mobile app usability were assessed, and evaluations were conducted on a scale of 1 through 5. Students who used Ethoshunt indicated whether they strongly agree or strongly disagree with the items that reflect their perceived level of positive emotions, ethical competency, user experience, and mobile app usability. Furthermore, students responded to five open-ended questions concerning Ethoshunt. The surveys were distributed to the students, and they returned the surveys to the researchers after completion.

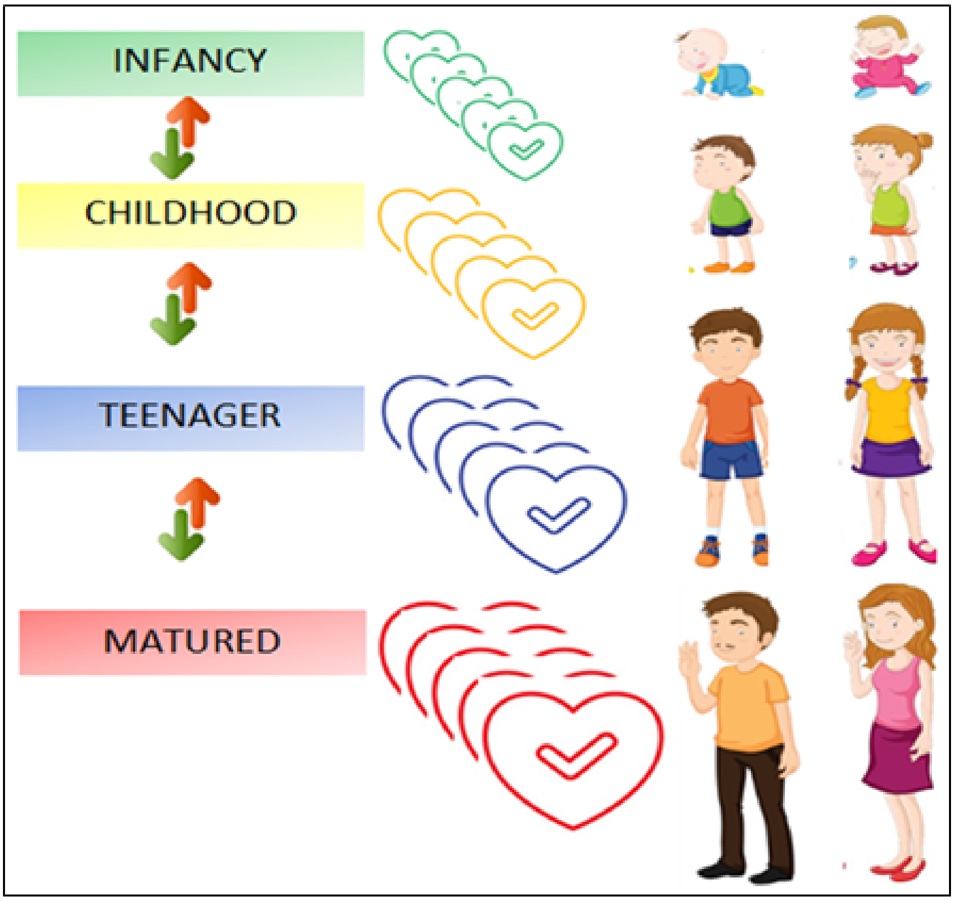
Results

Game Mechanics and Game Dynamics of Ethoshunt

Ethoshunt has been designed and created by the researchers through an understanding of game elements from traditional treasure hunt to aid learning, which is not seen as a serious game. Game mechanics refer to building blocks for gamifying offers, such as badges and scoring systems [35]. On the other hand, game dynamics describe the effects of game mechanics on students [35]. The game mechanics integrated in the mobile-based Ethoshunt trigger exploration in finding information relevant to the ethics course being taught, and there is a competition element as the game dynamic of Ethoshunt. Competition in Ethoshunt refers to a healthy game that creates collaboration among classmates or other game users in the long term. It involves a combination of game mechanics and motivational drivers to make Ethoshunt more fun and appealing to students. Levels, points, and rewards were also chosen as the game mechanics of Ethoshunt from other types of game mechanics, such as prizes, badges, and progress bars. In Ethoshunt, ethos levels eventually unlock the ethics course contents and ethos points increase the running numerical value of students' work.

Figure 1 presents the game plan of Ethoshunt for introducing four levels of ethics acquisition, which include comprehension and application. All students can choose the gender (male or female). They start at the infancy level and progress to the childhood level, teenager level, and eventually matured adult level. If they fail to secure points at any level, they may be demoted to an earlier level. Students who have achieved the matured adult level are considered competent, and they are perceived to have gained good understanding, experience, and application of the ethics course content. Ultimately, students at the matured adult level will experience a sense of pride in learning the ethics education course.

Figure 1. Game mechanics of Ethoshunt. The green arrow indicates advancement to the next level, and the red arrow indicates demotion to the previous level.

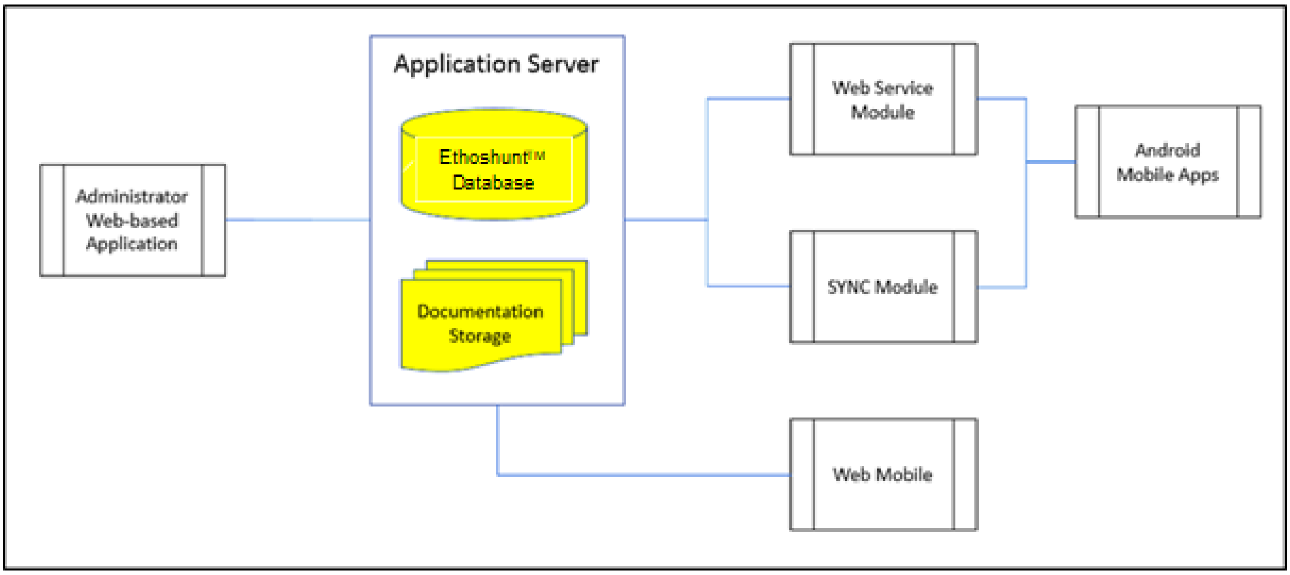


Ethoshunt System Architecture

Figure 2 illustrates the system architecture used in developing Ethoshunt. The system architecture was designed to handle three types of interfaces, including mobile web browser, desktop web browser, and Android mobile app. The administrators have a lot of information to view; hence, specifically for administrators, Ethoshunt is designed to be viewed using a desktop web browser. The system has been tested on the desktop web browsers Google Chrome (version 74.0.3729.131, official build, 64-bit; Google) on Windows 10 (Microsoft) and Safari (version 11.0.3; Apple) on Mac OS X El Capitan (version 10.11.6; Apple).

Users can use either a mobile web browser or the Android mobile app to access the system. The system has been tested on the mobile web browser Google Chrome (version 74.0.3729.157; Google), and the Android mobile app has been tested on Android Pie (version 9.0; Google).

Figure 2. Ethoshunt module workflow.



All views from the mobile web browser, desktop web browser, and Android mobile app are connected to the application server that stores the Ethoshunt database. A unique architecture needs to be implemented for the Android mobile app, where a web service module and sync module need to be placed in between the interface and server to establish a compatible connection between the app and the server. This ensures a real-time update of the Android mobile app once the server has been updated using either a desktop or mobile web browser.

Ethoshunt Functionality

Game mechanics and game dynamics are two crucial components in a gamification-based mobile app. Hint is used as a game mechanic in Ethoshunt. The ethics educator sends hints relevant to the ethics course contents to the students (Figure 3). It can be at any time of the day. The dedicated Android app receives the hints automatically via message, and an alert is displayed. The types of hints can lead to two types of activities as follows: (1) hidden information in the virtual world that requires students to find internet stories, links, movies, notes, quotes, images, and sounds related to the topics being discussed, with guidance by the hints and (2) hidden information in the real world that requires students to find physical clues related to the topics being discussed, with guidance by the hints. Students then submit the answers to the ethics educator and are given points for correct answers.

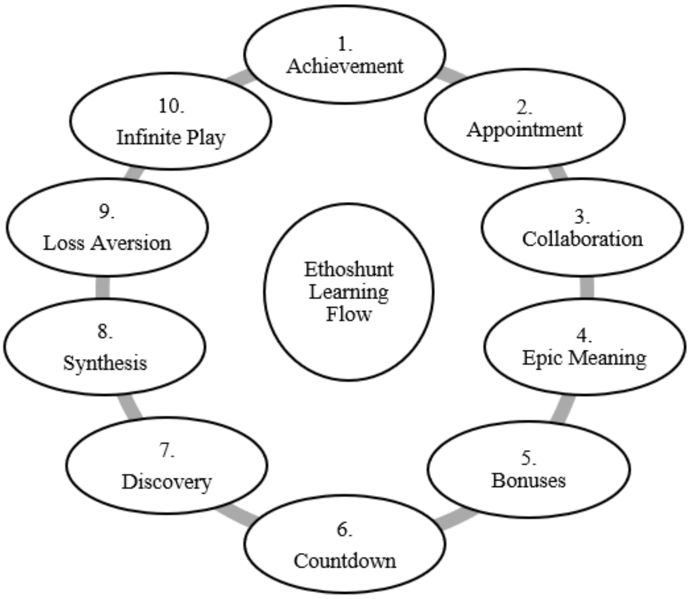
Figure 3. Hints sent by the ethics course instructor.

In the context of this study, a total of 39 students used Ethoshunt. The ethics educator initially sent a hint to the students after completion of the first three topics. They responded to the hint by submitting the required format of answers to the ethics educator through Ethoshunt. Points were given for correct answers, and the process was repeated throughout the semester. The achievement of each student, which is the level of ethics acquisition, was determined at the end of the semester. Examples of hints sent to the counseling students and types of hints required by the ethics educator are as follows: (1) What makes counseling supervision effective? (text); (2) What are the ethical dilemmas encountered in counseling research? (animation); (3) Therapeutic counseling relationship (image); (4) The counselor as a person and professional (video); (5) What a counselor should do when his or her client reports suicidal ideation? (text); (6) Songs relevant to confidentiality (audio).

Learning Flow Through Integrative Ethoshunt

Students who use Ethoshunt will be able to take part in the learning process through a meaningful educational flow. Through this flow, students develop their competence in ethics education and meaningfully reflect on what has been taught by the ethics educator prior to making their own revisions. Furthermore, this flow strengthens the process of conveying ethics course contents using gamification. Figure 4 shows the learning flow that every student will experience. First, achievement enables students to earn public recognition (learning peers and educator) for completing the tasks given. Second, students will check on their appointments, where they need to check in to receive new hints from their ethics educator. Third, students will collaborate and work with other learning peers to accomplish learning goals. Fourth, students are encouraged to work on the epic meaning, where they are expected to work to achieve excellent learning outcomes. Thereafter, students will receive unexpected rewards or bonuses for their achievements, and this is to motivate students to move forward and set countdowns. Through countdowns, students will tackle challenges in a limited amount of time. The seventh component in the flow is discovery. Students will navigate through the learning space and uncover pockets of knowledge. Students will then learn to synthesize the knowledge they have gained and work continuously on challenges that require multiple skills to solve. The ninth component is loss aversion, where students play to avoid losing what they have already gained, and finally, students will go through infinite play, where they continuously learn until becoming experts in the field. Students who are active throughout the learning process involving Ethoshunt will achieve unexpected positive outcomes.

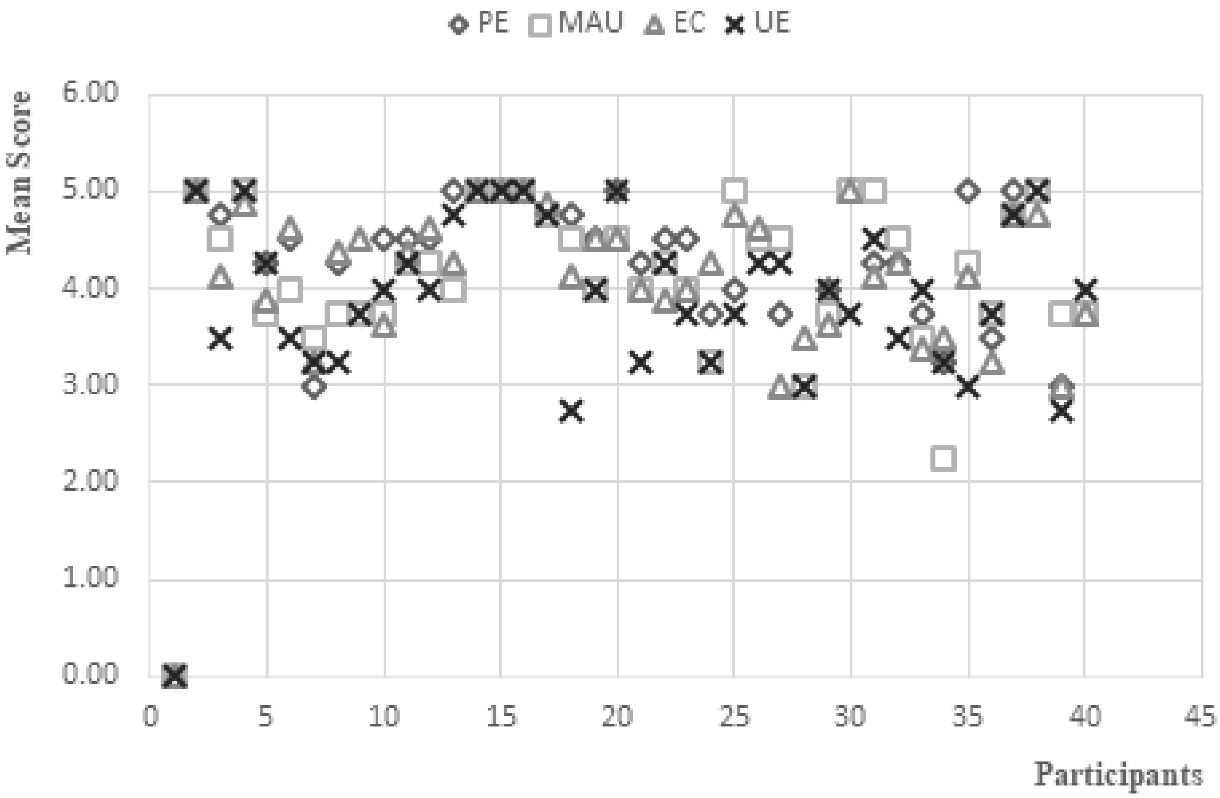
Figure 4. Learning flow through Ethoshunt.



Survey Analysis

The mean scores of positive emotions, ethical competency, user experience, and mobile app usability were calculated. The mean scores ranged from 2.2 to 5.0, which indicates that most of the participants' selections on the scale fell between 2 (disagree) and 5 (strongly agree). Figure 5 shows the distribution of the mean scores among the participants. The distribution pattern suggested that there were relevant relationships among positive emotions, ethical competency, user experience, and mobile app usability.

Figure 5. Distribution of mean scores. EC: ethical competency; MAU: mobile app usability; PE: positive emotions; UE: user experience.



For the first stage of analysis (Table 1), Pearson correlation indicated that positive emotions, ethical competency, and user experience were significantly related to mobile app usability. Positive emotions were positively related to mobile app usability (*r*=0.744, *P*<.001). There were also significant positive relationships between ethical competency and mobile app usability (*r*=0.686, *P*<.001) and user experience and mobile app usability (*r*=0.614, *P*<.001).

Table 1. Correlation coefficients for positive emotions, ethical competency, and user experience.

Variable/relationship	Correlation with mobile app usability (*r*)
Positive emotions	0.744
Ethical competency	0.686
User experience	0.614

Thereafter, Guildford rule of thumb was used to determine the strength of the relationships. Based on the rule of thumb (Multimedia Appendix 2), positive emotions had high correlation, followed by ethical competency and user experience, which had moderate correlation with mobile app usability.

Insights From Participants

For the second stage of analysis and referring to question one in Multimedia Appendix 1, it was found that 74% (29/39) of students had a better understanding of ethics education. They mentioned that their knowledge about ethics expanded, and it was their most meaningful experience. However, the opinions of 26% (10/39) of students were focused specifically on the ethics educator, ethics course, and classroom, instead of delivering perceptions about their most meaningful experiences in learning ethics.

For the second question, 36% (14/39) of students mentioned that the opportunity given to them to present in the class and the ethics educator's sharing of own experiences in ethics education were the most helpful experiences in the ethics class. They were able to make their own reflections based on the support provided by the ethics educator and were able to share their reflections with their classmates. One of the students mentioned that knowledge gained through the technology-based education tool was the most helpful experience, as the student could learn while enhancing the understanding of ethics. The remaining students stated that the activities conducted in the ethics class and the support from the ethics educator were the most helpful experiences in learning ethics.

In responding to what can be improved in the ethics education class, 26% (10/39) of students were expecting technology-based elements in learning ethics and 74% (29/39) of students focused on traditional teaching methods. For the fourth question, 38% (11/39) of students stated that the technology-mediated ethics education class can be improved by providing high-speed internet facilities in the classroom, extending the duration of the ethics class, and including more game elements in the mobile app. The rest of the students suggested frequent use of game-based tools in the classroom, as well as an attractive, colorful, and user-friendly mobile app. Finally, for the last question, 59% (23/39) of students wanted to see improvement in the elements of Ethoshunt. They suggested more game elements, more challenging questions, and more button functions, as well as the inclusion of animation with motions, attractive graphics, and ethics notes in games to improve ethics education classes that use Ethoshunt. The remaining students commented on the external factors of Ethoshunt, such as accessing Ethoshunt offline and internet speed.

Discussion

Principal Findings

In this study, the researchers found relationships between mobile app usability and positive emotions, ethical competency, and user experience. In addition, the researchers gathered new insights to be incorporated as ideas of gamification in ethics education. It is interesting to note that most of the students had meaningful experiences in learning ethics, and they considered Ethoshunt as a good mobile app with room for improvement.

The results of the survey analysis showed that there were relevant relationships between positive emotions and mobile app usability, ethical competency and mobile app usability, and user experience and mobile app usability. The highest correlation with mobile app usability was noted for positive emotions, followed by ethical competency and user experience. The results are in line with findings from other studies, which emphasize positive emotions as the effect of gamification techniques [35,36]. Blohm and Lelmelster [6] mentioned that gamification promotes behavior changes that could be the result of positive emotions exhibited after being exposed to gamification elements. Students who have experienced positive emotions as a result of using gamification in ethics education would most probably support mobile app usability. They would consider a gamification-based mobile app as a useful tool that can assist them in learning ethics. On the other hand, ethical competency and user experience recorded moderate relationships with mobile app usability. It can be seen that ethical competency and user experience do not greatly alter the idea of using gamification in ethics education. Although Ethoshunt was used to enhance the understanding of ethics education, students still perceived that there was not much difference in their ethical competency after using Ethoshunt. Their experience of using Ethoshunt was also almost similar to what they experience with traditional teaching and learning methods. Their positive emotions flourished more compared with ethical competency and user experience, which indicates the great impact of using gamification in ethics education. The findings reflect that positive emotions were associated with the usability of Ethoshunt. The primary objective of developing Ethoshunt was to improve students' ethical competency. However, the findings showed that students prefer a fun, contented, and interesting learning environment to learn ethics education. This would eventually help them to invest more time in exploring Ethoshunt and learn ethics education simultaneously.

The second stage of analysis that involved qualitative data reported various comments or feedback from students. Most of the students experienced fun and joy in learning ethics. They preferred more games in ethics education. However, they suggested providing high-speed internet and more time to explore Ethoshunt. All students participating in the study owned mobile phones and were able to access mobile data. However, the speed of the internet varies depending on multiple factors, such as network, mobile processor, and type of mobile phone. In addition, the design of the mobile app affects usability. A mobile app that is designed well will definitely attract users and increase the chance of improving usability [43].

Limitations and Future Research

Past literature reviews relevant to gamification that can explain positive emotions, ethical competency, user experience, and mobile app usability are very limited. This restricted the researchers in providing arguments that could support the research findings. Ethoshunt requires good internet connectivity, which is a constraint for some students who do not have access to smartphones and the internet. As for gamification itself, the newly developed online-based mobile app Ethoshunt can be used for all ethics-based courses, which can be expanded in the future to benefit other disciplines, such as ethics in engineering, ethics in medicine, ethics in law, ethics in business and marketing, ethics in technology, etc. In the future, a larger sample size can be used to broaden the perspectives of gamification, and assessments should be conducted in large-scale actual research.

Conclusion

Ethics education is one of the most important areas of knowledge acquisition in any profession. It is the responsibility of ethics educators to prepare students to learn, understand, experience, and apply ethics education to their personal and professional work. Lack of ethics knowledge and self-care awareness can contribute to poor ethical decision-making ability. Therefore, it is necessary to implement new strategies in teaching and learning to enhance the understanding of ethics education among students. The integration of gamification elements, integration of the use of game dynamics and game mechanics, and implementation of gamification deserve greater attention. Ethoshunt, a gamification-based mobile app, is one of the tools that can be integrated in teaching and learning ethics, as it enables creative and intuitive skills, which may encourage students to learn ethics more attractively and effectively. Furthermore, the nature of ethics education is complicated, and it is quite challenging to comprehensively teach ethics education in a semester. Thus, ethics educators may use Ethoshunt (although it requires improvement) to disseminate and apply ethics-related knowledge while making learning enjoyable.

Acknowledgments

The project has been sponsored by Universiti Putra Malaysia through a Teaching and Learning Incentive Grant for Gamification (grant number: 9323702).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Open-ended questions regarding Ethoshunt.
[DOCX File , 12 KB - games_v8i3e18247_app1.docx]

Multimedia Appendix 2
Guilford rule of thumb illustrating the strength of relationships.
[DOCX File , 12 KB - games_v8i3e18247_app2.docx]

References

1. Bovermann K, Bastiaens TJ. Towards a motivational design? Connecting gamification user types and online learning activities. RPTEL 2020 Jan 10;15(1). [doi: 10.1186/s41039-019-0121-4]
2. Rodrigues P, Souza M, Figueiredo E. Games and Gamification in Software Engineering Education: A Survey with Educators. 2019 Presented at: 2018 IEEE Frontiers in Education Conference (FIE); October 3-6, 2018; San Jose, CA, USA. [doi: 10.1109/fie.2018.8658524]
3. Tsay CH, Kofinas A, Luo J. Enhancing student learning experience with technology-mediated gamification: An empirical study. Computers & Education 2018 Jun;121:1-17. [doi: 10.1016/j.compedu.2018.01.009]
4. UNESCO ICT Competency Framework for Teachers. UNESCO. URL: https://unesdoc.unesco.org/ark:/48223/pf0000213475 [accessed 2018-06-28]
5. Garcia-Sanjuan F, Jurdi S, Jaen J, Nacher V. Evaluating a tactile and a tangible multi-tablet gamified quiz system for collaborative learning in primary education. Computers & Education 2018 Aug;123:65-84. [doi: 10.1016/j.compedu.2018.04.011]
6. Nousiainen T, Kangas M, Rikala J, Vesisenaho M. Teacher competencies in game-based pedagogy. Teaching and Teacher Education 2018 Aug;74:85-97. [doi: 10.1016/j.tate.2018.04.012]
7. Ferro LS. An analysis of players' personality type and preferences for game elements and mechanics. Entertainment Computing 2018 Aug;27:73-81. [doi: 10.1016/j.entcom.2018.03.003]
8. Hsu C, Chen M. How does gamification improve user experience? An empirical investigation on the antecedences and consequences of user experience and its mediating role. Technological Forecasting and Social Change 2018 Jul;132:118-129. [doi: 10.1016/j.techfore.2018.01.023]
9. Eppmann R, Bekk M, Klein K. Gameful Experience in Gamification: Construction and Validation of a Gameful Experience Scale [GAMEX]. Journal of Interactive Marketing 2018 Aug;43:98-115. [doi: 10.1016/j.intmar.2018.03.002]
10. Deterding S. Gamification. Interactions 2012 Jul 01;19(4):14. [doi: 10.1145/2212877.2212883]
11. van Benthem J. Computation as social agency: What, how and who. Information and Computation 2018 Aug;261:519-535. [doi: 10.1016/j.ic.2017.09.009]
12. Timpel P, Cesena FH, da Silva Costa C, Soldatelli MD, Gois E, Castrillon E, et al. Efficacy of gamification-based smartphone application for weight loss in overweight and obese adolescents: study protocol for a phase II randomized controlled trial. Ther Adv Endocrinol Metab 2018 Jun;9(6):167-176 [FREE Full text] [doi: 10.1177/2042018818770938] [Medline: 29854386]
13. Li C. Consumer behavior in switching between membership cards and mobile applications: The case of Starbucks. Computers in Human Behavior 2018 Jul;84:171-184. [doi: 10.1016/j.chb.2017.12.042]
14. Weber J, Azad M, Riggs W, Cherry CR. The convergence of smartphone apps, gamification and competition to increase cycling. Transportation Research Part F: Traffic Psychology and Behaviour 2018 Jul;56:333-343. [doi: 10.1016/j.trf.2018.04.025]
15. Aubert AH, Bauer R, Lienert J. A review of water-related serious games to specify use in environmental Multi-Criteria Decision Analysis. Environmental Modelling & Software 2018 Jul;105:64-78. [doi: 10.1016/j.envsoft.2018.03.023]
16. AlSkaif T, Lampropoulos I, van den Broek M, van Sark W. Gamification-based framework for engagement of residential customers in energy applications. Energy Research & Social Science 2018 Oct;44:187-195. [doi: 10.1016/j.erss.2018.04.043]
17. Zhang M, Ying J, Song G, Fung DS, Smith H. Gamified Cognitive Bias Modification Interventions for Psychiatric Disorders: Review. JMIR Ment Health 2018 Oct 25;5(4):e11640 [FREE Full text] [doi: 10.2196/11640] [Medline: 30361194]

18. Vélez Meza E, Alexis G, Mónica G, Jacinto M. Digital gamification in basic general education students. In: Basantes-Andrade A, Naranjo-Toro M, Zambrano Vizuete M, Botto-Tobar M, editors. Technology, Sustainability and Educational Innovation (TSIE). TSIE 2019. Advances in Intelligent Systems and Computing. Cham: Springer; 2020:143-156.
19. Markopoulos AP, Fragkou A, Kasidiaris PD, Davim JP. Gamification in engineering education and professional training. International Journal of Mechanical Engineering Education 2015 Jun 18;43(2):118-131. [doi: 10.1177/0306419015591324]
20. Negruşa A, Toader V, Sofică A, Tutunea M, Rus R. Exploring Gamification Techniques and Applications for Sustainable Tourism. Sustainability 2015 Aug 14;7(8):11160-11189. [doi: 10.3390/su70811160]
21. Alhammad MM, Moreno AM. Gamification in software engineering education: A systematic mapping. Journal of Systems and Software 2018 Jul;141:131-150. [doi: 10.1016/j.jss.2018.03.065]
22. Su C, Cheng C. A mobile gamification learning system for improving the learning motivation and achievements. Journal of Computer Assisted Learning 2014 Nov 06;31(3):268-286. [doi: 10.1111/jcal.12088]
23. Wiggins B. An overview and study on the use of games, simulations, and gamification in higher education. International Journal of Game-Based Learning 2016;6(1):18-29. [doi: 10.4018/ijgbl.2016010102]
24. Xi N, Hamari J. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. International Journal of Information Management 2019 Jun;46:210-221. [doi: 10.1016/j.ijinfomgt.2018.12.002]
25. Zhang M, Ying J, Ho RC. Gamified Cognitive Bias Interventions for Psychiatric Disorders: Protocol of a Systematic Review. JMIR Res Protoc 2018 Oct 16;7(10):e10154 [FREE Full text] [doi: 10.2196/10154] [Medline: 30327292]
26. Zakaria N, Warren J. Inquiry-Based Teaching and Learning in Counseling Ethics Education. In: Blessinger P, Carfora JM, editors. Inquiry-Based Learning for the Arts, Humanities, and Social Sciences: A Conceptual and Practical Resource for Educators (Innovations in Higher Education Teaching and Learning). Bingley, United Kingdom: Emerald Group Publishing Limited; 2014:147-167.
27. Chung C, Shen C, Qiu Y. Students' acceptance of gamification in higher education. International Journal of Game-Based Learning 2019;9(2):1-19. [doi: 10.4018/ijgbl.2019040101]
28. Davis K, Sridharan H, Koepke L, Singh S, Boiko R. Learning and engagement in a gamified course: Investigating the effects of student characteristics. J Comput Assist Learn 2018 Apr 19;34(5):492-503. [doi: 10.1111/jcal.12254]
29. Özdener N. Gamification for enhancing Web 2.0 based educational activities: The case of pre-service grade school teachers using educational Wiki pages. Telematics and Informatics 2018 Jun;35(3):564-578. [doi: 10.1016/j.tele.2017.04.003]
30. Ryan RM, Williams GC, Patrick H, Deci EL. Self-determination theory and physical activity: The dynamics of motivation in development and wellness. Hellenic Journal of Psychology 2009;6:107-124.
31. Charlier N, De Fraine B. Game-based learning in teacher education: A strategy to integrate digital games into secondary schools. International Journal of Game-Based Learning 2012;2(2). [doi: 10.4018/ijgbl.2012040101]
32. Zakaria N. Counseling ethics education experience: An interpretive case study of the first-year master's level counseling students (Doctoral dissertation). ProQuest Dissertations and Theses 2013 (UMI 3562075) 2013 May 11.
33. Fredrickson BL. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist 2001;56(3):218-226. [doi: 10.1037/0003-066x.56.3.218]
34. Sauter DA. The Nonverbal Communication of Positive Emotions: An Emotion Family Approach. Emot Rev 2017 Jul;9(3):222-234 [FREE Full text] [doi: 10.1177/1754073916667236] [Medline: 28804510]
35. Blohm I, Leimeister JM. Gamification: Design of IT-Based Enhancing Services for Motivational Support and Behavioral Change. Bus Inf Syst Eng 2013 Jun 14;5(4):275-278. [doi: 10.1007/s12599-013-0273-5]
36. Korn O, Boffo S, Schmidt A. The Effect of Gamification on Emotions - The Potential of Facial Recognition in Work Environments. In: Kurosu M, editor. Human-Computer Interaction: Design and Evaluation. HCI 2015. Lecture Notes in Computer Science. Cham: Springer; Jul 21, 2015:489-499.
37. Poikkeus T, Leino-Kilpi H, Katajisto J. Supporting ethical competence of nurses during recruitment and performance reviews - the role of the nurse leader. J Nurs Manag 2014 Sep;22(6):792-802. [doi: 10.1111/jonm.12033] [Medline: 23465074]
38. Barkhordari-Sharifabad M, Ashktorab T, Atashzadeh-Shoorideh F. Ethical competency of nurse leaders: A qualitative study. Nurs Ethics 2018 Feb;25(1):20-36. [doi: 10.1177/0969733016652125] [Medline: 27301416]
39. Zakaria N, Warren J. Counseling Ethics Education: Teaching and Learning Development Reformation. In: Amzat I, Yusuf B, editors. Fast forwarding Higher Education Institutions for Global Challenges. Singapore: Springer; 2016:83-96.
40. Zakaria N, Warren J, Bakar A. Counseling Ethics Education for Enhanced Professional Identity and Development: Guidance and Counseling Teachers Lifelong Learning Acquisition Empowered. In: Amzat I, Valdez N, editors. Teacher Empowerment Toward Professional Development and Practices. Singapore: Springer; 2017:153-166.
41. Sheng ML, Teo TS. Product attributes and brand equity in the mobile domain: The mediating role of customer experience. International Journal of Information Management 2012 Apr;32(2):139-146. [doi: 10.1016/j.ijinfomgt.2011.11.017]
42. He J, Chaparro BS, Haskins C. USee: A Mobile Usability Research Tool Using Google Glass. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2014 Oct 17;58(1):1242-1246. [doi: 10.1177/1541931214581259]
43. Wisniewski J. Mobile usability. Bul. Am. Soc. Info. Sci. Tech 2011 Dec 01;38(1):30-32. [doi: 10.1002/bult.2011.1720380110]

44. Jeong W, Han H. Usability study on mobile Web newspaper sites. Proc. Am. Soc. Info. Sci. Tech 2012 Jan 11;48(1):1-4. [doi: 10.1002/meet.2011.14504801287]
45. Akta Kaunselor 1998 (Akta 580) Dan Peraturan-peraturan: & Counsellors Act 1998 (Act 580) and Regulations. Malaysia: International Law Book Services; 2016.

Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by S Amat, J Willig; comments to author 10.03.20; revised version received 04.05.20; accepted 03.06.20; published 10.08.20.

Please cite as:
Zakaria NS, Saripan MI, Subarimaniam N, Ismail A
Assessing Ethoshunt as a Gamification-Based Mobile App in Ethics Education: Pilot Mixed-Methods Study
JMIR Serious Games 2020;8(3):e18247
URL: http://games.jmir.org/2020/3/e18247/
doi:10.2196/18247
PMID:32663153

Original Paper

Effect of Computer Debriefing on Acquisition and Retention of Learning After Screen-Based Simulation of Neonatal Resuscitation: Randomized Controlled Trial

Daphne Michelet[1,2], MD; Jessy Barre[1], PhD; Jennifer Truchot[1,3], PhD; Marie-Aude Piot[1,4], MD; Philippe Cabon[1,5], PhD; Antoine Tesniere[1,6], Prof Dr

[1]Ilumens Platform of Simulation in Healthcare, Université de Paris, Paris, France

[2]Department of Anesthesia and Intensive Care, American Memorial Hospital, Centre Hospitalier Universitaire de Reims, Reims, France

[3]Emergency Department, Lariboisière University Hospital, Paris, France

[4]Psychiatry Department, Monsouris Mutualiste Institute, Paris, France

[5]Laboratoire Adaptation Travail Individu, Université de Paris, Boulogne-Billancourt, France

[6]Department of Anesthesia and Intensive Care, Georges Pompidou European Hospital, Paris, France

Corresponding Author:
Daphne Michelet, MD
Ilumens Platform of Simulation in Healthcare
Université de Paris
45 rue des Saints Pères
Paris, 75006
France
Phone: 33 326832537
Email: daphnemichelet@gmail.com

Abstract

Background: Debriefing is key in a simulation learning process.

Objective: This study focuses on the impact of computer debriefing on learning acquisition and retention after a screen-based simulation training on neonatal resuscitation designed for midwifery students.

Methods: Midwifery students participated in 2 screen-based simulation sessions, separated by 2 months, session 1 and session 2. They were randomized in 2 groups. Participants of the debriefing group underwent a computer debriefing focusing on technical skills and nontechnical skills at the end of each scenario, while the control group received no debriefing. In session 1, students participated in 2 scenarios of screen-based simulation on neonatal resuscitation. During session 2, the students participated in a third scenario. The 3 scenarios had an increasing level of difficulty, with the first representing the baseline level. Assessments included a knowledge questionnaire on neonatal resuscitation, a self-efficacy rating, and expert evaluation of technical skills as per the Neonatal Resuscitation Performance Evaluation (NRPE) score and of nontechnical skills as per the Anaesthetists' Non-Technical Skills (ANTS) system. We compared the results of the groups using the Mann-Whitney U test.

Results: A total of 28 midwifery students participated in the study. The participants from the debriefing group reached higher ANTS scores than those from the control group during session 1 (13.25 vs 9; U=47.5; *P*=.02). Their scores remained higher, without statistical difference during session 2 (10 vs 7.75; *P*=.08). The debriefing group had higher self-efficacy ratings at session 2 (3 vs 2; U=52; *P*=.02). When comparing the knowledge questionnaires, the significant baseline difference (13 for debriefing group vs 14.5 for control group, *P*=.05) disappeared at the end of session 1 and in session 2. No difference was found for the assessment of technical skills between the groups or between sessions.

Conclusions: Computer debriefing seems to improve nontechnical skills, self-efficacy, and knowledge when compared to the absence of debriefing during a screen-based simulation. This study confirms the importance of debriefing after screen-based simulation.

Trial Registration: ClinicalTrials.gov NCT03844009; https://clinicaltrials.gov/ct2/show/NCT03844009

(JMIR Serious Games 2020;8(3):e18633) doi:10.2196/18633

KEYWORDS

screen-based simulation; debriefing; neonatal resuscitation; simulation; medical education; midwifery; neonatal

Introduction

Neonatal resuscitation requires training. Almost 10% of newborns and 80% of preterm newborns weighing less than 1500 g will undergo resuscitation at birth, and the quality of care provided during the first minute of life is directly linked to the outcome [1-3]. Theoretical knowledge from current guidelines [4] is essential to ensure optimal neonatal resuscitation. Several technical skills, such as bag-mask ventilation, endotracheal intubation, or umbilical catheter placement, and nontechnical skills, such as situation awareness, decision making, communication, and teamwork [4,5] are also required to ensure safety and efficacy.

Since 2011, the Neonatal Resuscitation Program (NRP) developed by the American Academy of Pediatrics includes simulation-based training. The implementation of NRP led to a decrease in neonatal and perinatal mortality [6]. Simulation training increases the trainees' self-confidence [7], knowledge [2], and technical skills [8] and improves team behavior [9]. Simulation training has many advantages such as the possibility to practice procedures without any risk for the patient and for trainees to commit errors and learn from those errors, through the repetition of different scenarios [9].

In recent years, screen-based simulation has become increasingly prevalent. They show many advantages such as better affordability than high-fidelity simulation [10], transportable, and autonomous (no need for an instructor). Screen-based simulation appears to be a valid tool in simulation-based education for health professionals, ensuring the same learning efficacy than traditional learning methods [11,12]. Indeed, recently, the development of computer sciences allowed the creation of more realistic medical simulators to improve knowledge and acquire nontechnical skills, know-how, and technical gestures [12,13]. A screen-based simulator (NRP eSim) designed by Laerdal Medical in collaboration with the American Academy of Pediatrics is even included in the NRP program as 1 of the 6 educational components of the NRP 7th edition curriculum [14].

Debriefing is inseparable from simulation. It has been shown to improve professional practice and clinical skills [15-18]. Debriefing represents a discussion between 2 or more individuals during which, aspects of a performance are explored and analyzed with the aim of gaining insights that impact the quality of future clinical practice [19]. Various efficient debriefing methods exist: postsimulation debriefing, in-simulation debriefing, verbal instructor debriefing, video-assisted instructor debriefing, self-debriefing, and multimedia debriefing (a computer text presentation with audio voice-over and videos) [15,17]. For example, Boet et al [20] showed that a self-debriefing (formative self-assessment aiming to provide feedback, allowing students to reflect on their performance and subsequently improve their skills) was as effective as traditional debriefing by an instructor. As part of simulation-based education, screen-based simulation must provide debriefing. These simulators "can easily include tools and modules of various kinds to collect data transparently during play. The data can then be processed to provide material for feedback during play, as in-game debriefing, and also as part of the end-of-game debriefing" [21]. For example, after the NRP eSim training, students received automated feedback for self-reflection. This feedback highlighted good performances achieved during the experience, the procedures that needed to be improved, and the missed procedures. The feedback represents what we refer to as "computer debriefing," often delivered after a screen-based simulation in order to stay in a virtual environment with no need of an instructor [22]. However, few evaluations of the impact of computer debriefing on acquisition and retention of learning exist.

Retention of learning has been studied extensively after different simulation training in health sciences [23] and neonatal resuscitation [2,24,25]. However, the mean retention time of learning after simulation training and the optimal time interval between two formations remain debated [26,27]. The role of debriefing on retention of learning was already highlighted in some high-fidelity simulation studies [15,17], but it has not been studied in the context of screen-based simulation.

The objective of this study is to evaluate the impact of a computer debriefing after a screen-based simulation session compared to no debriefing in a virtual environment with no instructor. Our endpoints are acquisition of knowledge and skills and their retention after 2 months. We hypothesized that the debriefing group would yield better scores in different evaluations (knowledge, technical skills, nontechnical skills, and self-efficacy) as compared to the control group.

Methods

This randomized controlled simulation study was performed from November 2018 to January 2019 at L'école de Sages-Femmes de Baudelocque, a midwifery school of the Université de Paris. It was approved by the CERAR (Comité Ethique sur la Recherche en Anesthésie Réanimation) (IRB 00010254-2017-008). All students signed an informed written consent. The study was registered at ClinicalTrials.gov (NCT03844009).

Participants

Volunteer participants were recruited from among fourth-year students of L'école de Sages-Femmes de Baudelocque in Paris. They all followed the same curriculum on neonatal resuscitation, corresponding to only 1 academic course. No sample size calculation was performed for this research; a convenience sample was used. We included all 28 volunteers of the fourth-year class of 35 students.

Screen-Based Simulation

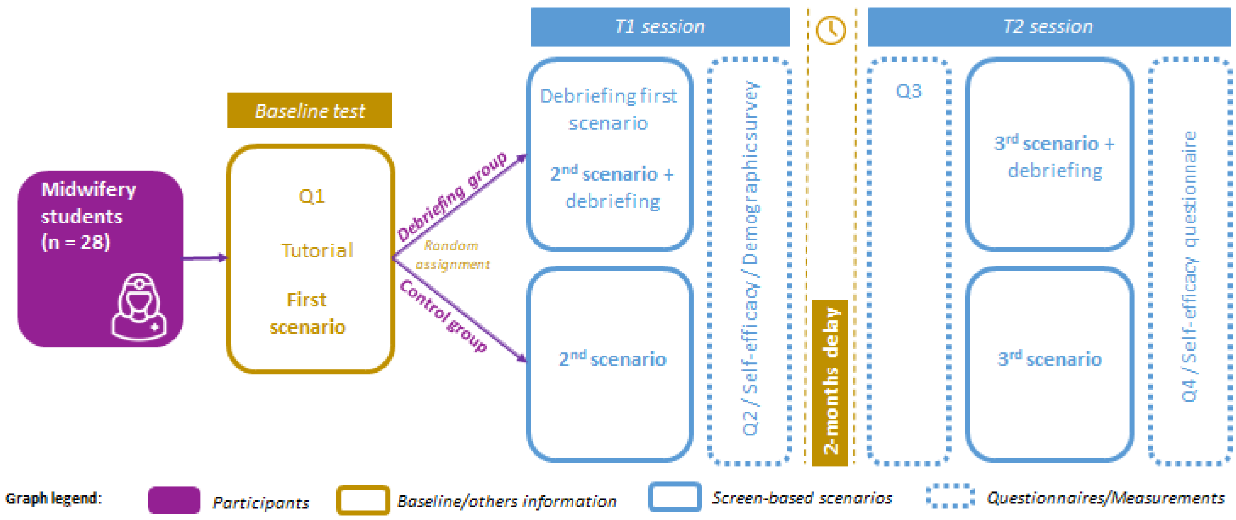
The screen-based simulation—Périnatsims—was designed by Medusims. It features the virtual environment of a delivery room in 3D with a newborn installed on a neonatal resuscitation table (Figure 1). The simulation used a point-and-click interface with a first-person point of view. In this digital simulator, learners could either be midwife, anesthetist, or pediatrician, although all the participants of this study were midwifery students. Throughout the scenario, the learner can call a pediatrician for help. Many scenarios were available with different difficulty levels (eg, preterm birth, emergency cesarean under general anesthesia, and abruptio placentae).

Figure 1. Participant during a scenario on the left and screenshot of the interface and virtual environment of Périnatsims screen-based simulation on the right.

Design

Each participant performed individually on a laptop during 2 screen-based simulation sessions: session 1 in November 2018 and session 2 in January 2019 (Figure 1). The study design is summarized in Figure 2.

Figure 2. Experiment design.



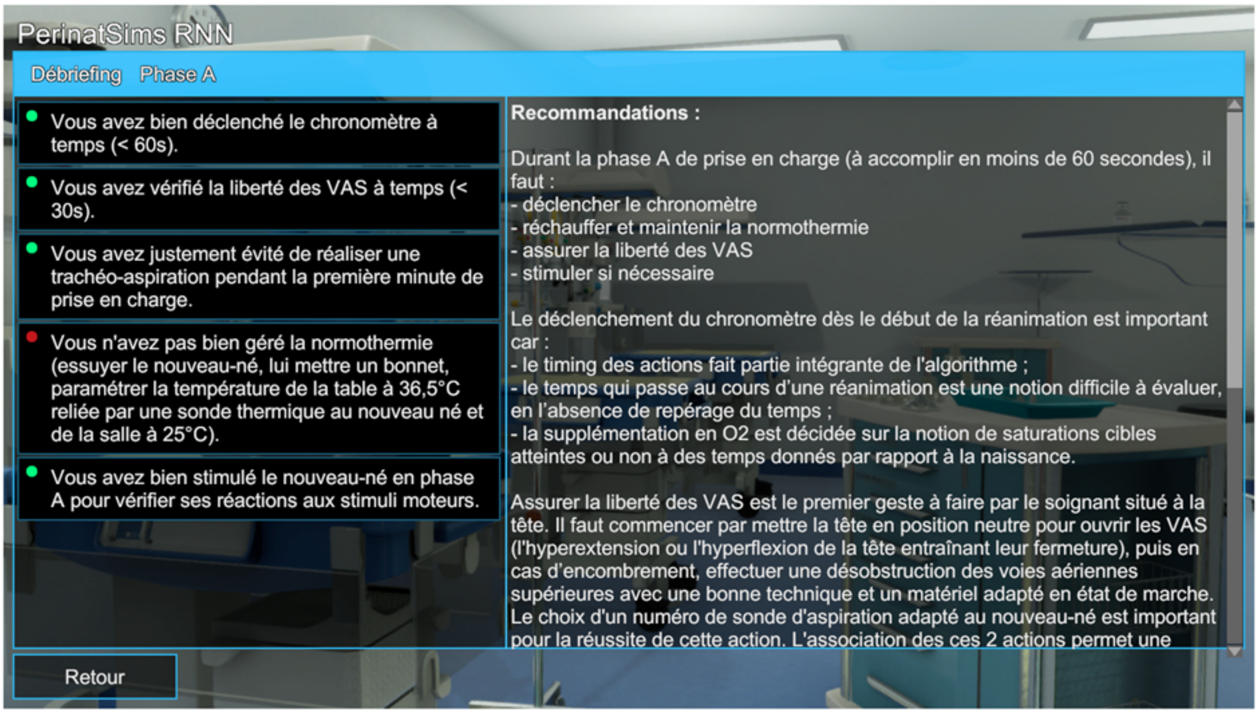
Session 1 started with a knowledge questionnaire (Q1, baseline), followed by the briefing, consisting of a 15-minute tutorial to explain different possible actions in the screen-based simulation. The participant performed the first scenario (low difficulty level) considered as the baseline level for knowledge and skills. It was followed by a second scenario (medium difficulty level). At the end of session 1, each participant again filled out the knowledge questionnaire (Q2) and a demographic survey, including a self-efficacy question.

Session 2 was conducted after 2 months. The simulation started with the same knowledge questionnaire (Q3) and tutorial, followed by a third scenario (high difficulty level). At the end of session 2, a last knowledge questionnaire (Q4) and the self-efficacy question were administered.

All 3 scenarios were identical and in the same order of increasing difficulty for every participant. The potential exposure of each participant to a real case (or training) of neonatal resuscitation during the 2 months delay was controlled and monitored.

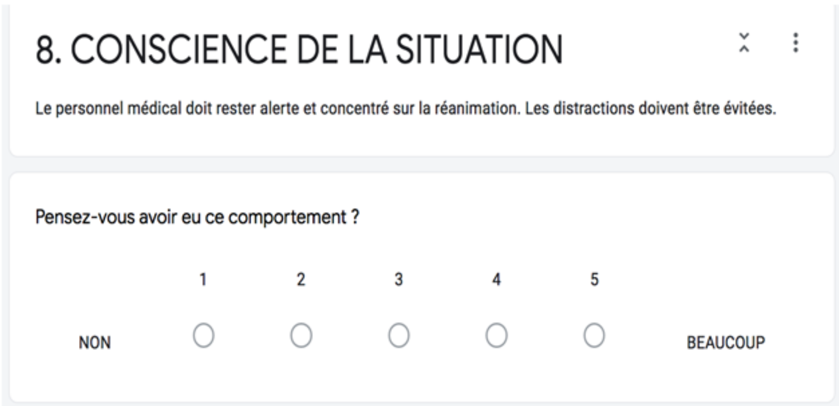
Participants were randomized in 2 groups: debriefing group and control group. At the end of each scenario, participants from the debriefing group accessed a computer debriefing on technical and nontechnical skills. Technical skills assessment, based on the recommendations of International Liaison Committee on Resuscitation (ILCOR), was presented with a color code: green (well-performed action), orange (partially-performed action), and red (absent or wrong action) (Figure 3).

Figure 3. Computer debriefing of technical skills.



Debriefing of the nontechnical skills was a self-debriefing. Each nontechnical skill involved in the neonatal resuscitation [3-5,28] was explained in one sentence, and then the learner self-rated their proficiency on a scale of 1 to 5, as shown in Figure 4. In this example, the nontechnical skill of situational awareness is explained as, “Medical staff have to stay alert and focus on the resuscitation. Distractions must be avoided.” The following question is, “Do you think you had this behavior?”

Figure 4. Self-debriefing of nontechnical skills.



Participants from the control group had no debriefing until the end of the second session and the completion of every questionnaire. The sessions were recorded using a camera with participants’ written consent.

Outcomes

Comparison of Knowledge Acquisition and Retention

Knowledge was assessed using validated questionnaires (25 questions with single or multiple choices) based on the ILCOR recommendations [29]: Q1 at the beginning of session 1, Q2 at the end of session 1, Q3 at the beginning of session 2, and Q4 at the end of session 2.

Comparison of Skills Acquisition and Retention

Two independent blinded raters (an anesthetist and a human factors expert specialized in health sciences) evaluated the technical and nontechnical skills retrospectively by analyzing the video recordings.

Technical skills were assessed by the Neonatal Resuscitation Performance Evaluation (NRPE) scoring system [30], with 20 points for each scenario (eg, checked the material, dried the newborn, and initiated mask ventilation). Nontechnical skills were assessed by the Anaesthetists’ Non-Technical Skills (ANTS) [31] observation system, including four categories: situation awareness, task management, team work, and decision making (eg, prioritizing, coordinating activities with team members, gathering information, and selecting options). ANTS is a validated tool used to assess nontechnical skills in various situations, ranging from emergencies for medical students [32] to neonatal resuscitation for midwives (with a specific modified ANTS version) [33]. ANTS scores were recorded as the overall category scores on a scale of 1-4 from poor performance to good performance and on the 16-point global score as per ANTS system. Interrater reliability calculations were performed for both evaluations, with a good agreement between the two raters (κ=0.66; *P*=.01).

Comparison of Self-Efficacy Evaluation

The self-efficacy question assessed midwives' perception on their own performance: "How much are you confident in your capability to organize and execute a neonatal resuscitation?" using a 6-point Likert scale ranging from "not at all confident" (scored as 0) to "very confident" (scored as 5) [34]. It was assessed at the end of each session.

Statistical Analysis

Data are presented as median (IQR) for continuous data given the small sample size. Agreement between raters for the ANTS and NRPE scores was evaluated using percent agreement and corresponding Cohen kappa coefficient (inter-rater agreement). Comparisons between groups were performed using the Mann-Whitney *U* test for independent samples. All tests were two-tailed, and statistical significance was considered at *P*<.05. Statistical analyses were performed using SPSS 25.0 software (IBM Corp).

Results

The study included 28 participants; 14 were randomly assigned to the control group and 14, to the debriefing group. The participants were fourth-year students of a 5-year curriculum of midwifery in France. A majority (27/28) were women. The median (IQR) age was 22 (21-22) years. Five participants had previously followed high-fidelity simulation training, one had followed screen-based simulation training, but none had followed any training on neonatal resuscitation. No participant had witnessed or participated in a real neonatal resuscitation in 2 months prior to the study or had received any training.

Comparison of Knowledge Acquisition and Retention

At baseline, the control group (median 14.5; IQR 12.5-16) had better results than the debriefing group (median 12.5; IQR 11-13.75) (*P*=.05). This difference disappeared over time. There is no difference between the groups during session 1 and session 2. Results are presented in Table 1.

Table 1. Comparison of the knowledge questionnaires of the control and debriefing groups.

Questionnaires of the groups	Median (IQR)	*U*	*P* value
Q1 baseline		**56**	**.05**
Debriefing group	12.5 (11-13.75)		
Control group	14.5 (12.5-16)		
Q2 at the end of session 1		**66**	**.15**
Debriefing group	13 (12.25-14)		
Control group	14 (12.25-16)		
Q3 in session 2		**82**	**.47**
Debriefing group	14.5 (13-16)		
Control group	14 (13.25-15)		
Q4 at the end of session 2		**83**	**.48**
Debriefing group	14 (13.25-14.75)		
Control group	14 (12.5-15.75)		

Comparison of Skills Acquisition and Retention

Technical Skills Assessment Through the NRPE

No significant difference in the NRPE scores was observed during the experimentation (Table 2).

Table 2. Comparison of the nontechnical skills, technical skills, and self-efficacy evaluation between the debriefing and control groups.

	Baseline				Session 1				Session 2			
	Debriefing	Control	*U*	*P* value	Debriefing	Control	*U*	*P* value	Debriefing	Control	*U*	*P* value
ANTS[a] score (total=16 points), median (IQR)	**8 (6.1-9.8)**	**6.75 (5.6-7.9)**	**78.5**	**.38**	**13.25 (11.1-14.4)**	**9 (6.6-11.4)**	**47.5**	**.02**	**10 (9.3-13.9)**	**7.75 (6.5-12)**	**60.5**	**.08**
Task management (total=4 points)	2 (1.37-2.5)	1.5 (1-2)	70	.18	3 (2.5-3.62)	2 (1.5-3)	43	.10	2 (2-3.62)	2 (1-3)	78	.34
Team work (total=4 points)	2 (1.37-3)	2 (1-2.5)	80	.39	3.75 (2.87-4)	2.25 (1.87-3.62)	54	.04	4 (3-4)	3.25 (1.87-4)	65.5	.11
Situation awareness (total=4 points)	2 (1.5-2.5)	2 (1.37-2.5)	87	.60	3.25 (2.87-3.62)	2.25 (1.5-3.62)	60	.08	2.25 (1.87-3.62)	1.5 (1-3)	65.5	.13
Decision making (total=4 points)	1.5 (1-2.5)	1.5 (1-2.5)	94	.85	3 (2.37-3.5)	2.25 (1.37-3.5)	61	.08	2 (1.87-3.5)	2 (1-3)	70	.19
NRPE[b] score (total=20 points), median (IQR)	10 (7.3-12.3)	10 (7.3-12.3)	94.5	.87	9.2 (7.7-13.5)	10 (9.2-13.1)	87.5	.62	10.5 (8-12.8)	10.5 (7.5-12)	97	.96
Self-efficacy (total=5 points), median (IQR)	N/A	N/A	N/A	N/A	2 (1-2)	2 (1-2)	92	.76	3 (2-3)	2 (1-2)	52	.02

[a]ANTS: Anaesthetists' Non-Technical Skills.

[b]NRPE: Neonatal Resuscitation Performance Evaluation.

Nontechnical Skills Assessment Through the ANTS

A significant difference was observed between the two groups for session 1 (*U*=47.5; *P*=.02) and remained higher in favor of the debriefing group during session 2 (*U*=60.5; *P*=.08), while no difference was found in the baseline evaluation (scenario 1). The results (including the subcategories analysis) are presented in Table 2.

Comparison of Self-Efficacy Evaluation

A significant difference was found between the groups for session 2, with an improved self-efficacy score for the debriefing group (Table 2).

Discussion

Major Findings

This study highlights the benefit of a computer debriefing compared to no debriefing on nontechnical skills acquisition, self-efficacy, and knowledge after a screen-based simulation of neonatal resuscitation. Our hypothesis that the debriefing group would obtain better scores than the control group in the different evaluations is validated for knowledge, nontechnical skills, and self-efficacy.

The major interest of debriefing after a simulation session has already been extensively demonstrated. The review by Cheng et al [15], including 108 studies comparing debriefing and no debriefing, found positive effects of debriefing on every knowledge and skills outcomes. From debriefing comes a major part of the theoretical benefit of screen-based simulation for training, contributing to meaningful connections between the learning experience and the real world [21]. However, our study was the first to compare computer debriefing and no debriefing and to analyze their impact on knowledge, technical skills, and nontechnical skills after a screen-based simulation.

Concerning knowledge evaluation, participants of the control group had better baseline knowledge of neonatal resuscitation than the debriefing group. Our results showed an improvement in the debriefing group's score from the baseline level. The differences between the groups disappeared at the end of sessions, reflecting a positive effect of debriefing.

Usually, personalized debriefing after screen-based simulation addresses only technical skills. Data collected from the simulation are given back in the form of an automated feedback at the end of the scenario [21]. We found no evolution for the technical skills in our study. However, the increasing difficulty of the scenarios was designed to minimize the repetition effect on performance, as repeating the same scenario increases the participants' skills more than varying the scenarios [35]. This could mask the effect of the debriefing itself since the required technical skills evolved with scenarios.

In this study, we added a self-debriefing of nontechnical skills after the screen-based simulation of a neonatal resuscitation. In a review on screen-based simulation for medical education and surgical skills training, Graafland et al [36] highlighted the interest of a screen-based simulation to train nontechnical skills. Furthermore, in a review of debriefing techniques after nontechnical skills simulation training, performance seemed to improve equally with various methods of debriefing: skilled facilitator, novice instructor using a script, and self-led debrief or multimedia debriefing [37]. Our results confirm the possibility and benefit of a self-debriefing of nontechnical skills following a screen-based simulation to improve learning.

The second major finding of this study is the effect of computer debriefing on retention 2 months after the initial training. The debriefing group showed a better self-efficacy assessment than the control group. Their ANTS performance remained higher than that of the control group. The role of debriefing on retention of learning was already underlined in some studies [25,38]. Few studies assessed the retention of learning after screen-based simulation training. Their results were rather positive when evaluated up to 1 month after simulation [38] but less effective than traditional learning methods when evaluated after 6 months [39]. Our positives results are encouraging and emphasize the role of the debriefing in retention of learning even though further studies are needed to confirm a longer-term effect.

Limitations of the Study

First, this study compared the effect of a computer debriefing with the effect of the absence of debriefing. Our objective was to stay in a virtual environment without the need for an instructor. As debriefing is a major component of simulation training, participants from the control group had access to the complete debriefing at the end of session 2. Therefore, this study only assessed the efficacy of a computer debriefing but not the superiority over other debriefing methods.

Second, the timing of the debriefing was not standardized or assessed. Participants from the debriefing group had an unlimited amount of time to consult the debriefing. This was not the case for the control group. Perhaps, a free time period should have also been proposed to the control group to offer the possibility for a spontaneous reflective process.

Third, the nontechnical skills assessment was performed with the ANTS scoring tool, which was not originally developed and validated for the studied population. The lack of published data on the use of the ANTS scores for midwives is a limitation.

Conclusion

Computer debriefing seems to improve nontechnical skills and self-efficacy estimation when compared to the absence of debriefing during a screen-based simulation. It also allows a progression of learner’s knowledge. This study supports the benefit of debriefing (including a computer debriefing) in screen-based simulation.

Acknowledgments

Authors would like to thank L’école de Sages-Femmes de Baudelocque, Université de Paris and especially, Ms Christele Vérot and all the students who participated in the study.

Authors' Contributions

DM participated in study design, student recruitment, simulation sessions, a posteriori analysis of the simulation sessions, data analysis and interpretation, drafting and revising the manuscript, and approved the final version. JB participated in study design, student recruitment, simulation sessions, a posteriori analysis of the simulation sessions, data analysis and interpretation, drafting and revising the manuscript, and approved the final version. JT participated in study design, a posteriori analysis of the simulation sessions, data interpretation, drafting and revising the manuscript, and approved the final version. MAP participated in a posteriori analysis of the simulation sessions and approved the final version. PC participated in data collection analysis and interpretation, drafting and revising the manuscript, and approved the final version. AT participated in data collection analysis and interpretation, drafting and revising the manuscript, and approved the final version.

Conflicts of Interest

None declared.

References

1. Ryan C, Clark L, Malone A, Ahmed S. The effect of a structured neonatal resuscitation program on delivery room practices. Neonatal Netw 1999;18:25 [FREE Full text] [doi: 10.1891/0730-0832.18.1.25] [Medline: 10205469]
2. Curran V, Aziz K, O?Young S, Bessell C. Evaluation of the effect of a computerized training simulator (ANAKIN) on the retention of neonatal resuscitation skills. Teach Learn Med 2004;16:157 [FREE Full text] [doi: 10.1207/s15328015tlm1602_7] [Medline: 15276892]
3. Pearlman S, Zern S, Blackson T, Ciarlo J, Mackley A, Locke R. Use of neonatal simulation models to assess competency in bag-mask ventilation. J Perinatol 2016;36:242 [FREE Full text] [doi: 10.1038/jp.2015.175] [Medline: 26583941]
4. Thomas E, Taggart B, Crandell S, Lasky R, Williams A, Love L. Teaching teamwork during the Neonatal Resuscitation Program: a randomized trial. J Perinatol 2007;27:a [FREE Full text] [doi: 10.1038/sj.jp.7211771] [Medline: 17538634]
5. Thomas E, Williams A, Reichman E, Lasky R, Crandell S, Taggart W. Team training in the neonatal resuscitation program for interns: teamwork and quality of resuscitations. Pediatrics 2010;125:539 [FREE Full text] [doi: 10.1542/peds.2009-1635] [Medline: 20156896]

6. Patel A, Khatib M, Kurhe K, Bhargava S, Bang A. Impact of neonatal resuscitation trainings on neonatal and perinatal mortality: a systematic review and meta-analysis. BMJ Paediatr Open 2017;1:a [FREE Full text] [doi: 10.1136/bmjpo-2017-000183] [Medline: 29637172]
7. Malmström B, Nohlert E, Ewald U, Widarsson M. Simulation-based team training improved the self-assessed ability of physicians, nurses and midwives to perform neonatal resuscitation. Acta Paediatr 1992:1273 [FREE Full text] [doi: 10.1111/apa.13861] [Medline: 28370414]
8. Lee M, Brown L, Bender J, Machan J, Overly F. A medical simulation-based educational intervention for emergency medicine residents in neonatal resuscitation. Acad Emerg Med 2012 May;19(5):577-585 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01361.x] [Medline: 22594362]
9. Mileder L, Urlesberger B, Szyld E, Roehr C, Schmölzer G. Simulation-based neonatal and infant resuscitation teaching: a systematic review of randomized controlled trials. Klin Padiatr 2014;226:a [FREE Full text] [doi: 10.1055/s-0034-1372621] [Medline: 25153910]
10. Haerling K. Cost-Utility Analysis of Virtual and Mannequin-Based Simulation. Simul Healthc 2018;13:33 [FREE Full text] [doi: 10.1097/sih.0000000000000280] [Medline: 29373382]
11. Gentry S, Gauthier A, L'Estrade EB, Wortley D, Lilienthal A, Tudor CL. Serious Gaming and Gamification Education in Health Professionsystematic Review. J Med Internet Res 2019;21:e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
12. Barré J, Michelet D, Job A, Truchot J, Cabon P, Delgoulet C. Does Repeated Exposure to Critical Situations in a Screen-Based Simulation Improve the Self-Assessment of Non-Technical Skills in Postpartum Hemorrhage Management? Simul Gaming 2019. Simulation and gaming 2019:1046878119827324 [FREE Full text] [doi: 10.1177/1046878119827324]
13. Ma M, Jain L, Anderson P. Virtual, augmented reality and serious games for healthcare 1. In: vol. 1. Springer. Verlag Berlin Heidelberg: Springer; 2014.
14. Sawyer T, Umoren R, Gray M. Neonatal resuscitation: advances in training and practice. Adv Med Educ Pract 2017;8:a [FREE Full text] [doi: 10.2147/amep.s109099] [Medline: 28096704]
15. Cheng A, Eppich W, Grant V, Sherbino J, Zendejas B, Cook D. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. Med Educ 2014;48:a [FREE Full text] [doi: 10.1111/medu.12432] [Medline: 24909527]
16. Sawyer T, Eppich W, Brett-Fleegler M, Grant V, Cheng A. More Than One Way to Debrief: A Critical Review of Healthcare Simulation Debriefing Methods. Simul Healthc 2016;11:A [FREE Full text] [doi: 10.1097/sih.0000000000000148] [Medline: 27254527]
17. Levett-Jones T, Lapkin S. A systematic review of the effectiveness of simulation debriefing in health professional education. Nurse Educ Today 2014;34:e58-e63 [FREE Full text] [doi: 10.1016/j.nedt.2013.09.020] [Medline: 24169444]
18. Kolbe M, Grande B, Spahn D. Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. Best Pract Res Clin Anaesthesiol 2015;29:87 [FREE Full text] [doi: 10.1016/j.bpa.2015.01.002] [Medline: 25902470]
19. Chronister C, Brown D. Comparison of Simulation Debriefing Methods. Clin Simul Nurs 2012;8:e281 [FREE Full text] [doi: 10.1016/j.ecns.2010.12.005]
20. Boet S, Bould M, Bruppacher H, Desjardins F, Chandra D, Naik V. Looking in the mirrorlf-debriefing versus instructor debriefing for simulated crises. Crit Care Med 2011;39:1377 [FREE Full text] [doi: 10.1097/ccm.0b013e31820eb8be] [Medline: 21317645]
21. Crookall D. Serious Games, Debriefing, and Simulation/Gaming as a Discipline. Simulation & Gaming 2011 Jan 06;41(6):898-920. [doi: 10.1177/1046878110390784]
22. Pasquier P, Mérat S, Malgras B, Petit L, Queran X, Bay C, et al. A Serious Game for Massive Training and Assessment of French Soldiers Involved in Forward Combat Casualty Care (3D-SC1): Development and Deployment. JMIR Serious Games 2016 May 18;4(1):e5 [FREE Full text] [doi: 10.2196/games.5340] [Medline: 27194369]
23. Roy K, Miller M, Schmidt K, Sagy M. Pediatric residents experience a significant decline in their response capabilities to simulated life-threatening events as their training frequency in cardiopulmonary resuscitation decreases. Pediatr Crit Care Med 2011;12:e141-e144 [FREE Full text] [doi: 10.1097/pcc.0b013e3181f3a0d1] [Medline: 20921919]
24. Kaczorowski J, Levitt C, Hammond M, Outerbridge E, Grad R, Rothman A. Retention of neonatal resuscitation skills and knowledge: a randomized controlled trial. Fam Med 1998;30:11. [Medline: 9827341]
25. Kudreviciene A, Nadisauskiene R, Tameliene R, Tamelis A, Nedzelskiene I, Dobozinskas P. Initial neonatal resuscitationkill retention after the implementation of the novel 24/7 HybridLab® learning system. J Matern-Fetal Neonatal Med 2019;32:1230 [FREE Full text] [doi: 10.1080/14767058.2017.1402881] [Medline: 29117772]
26. Au K, Lam D, Garg N, Chau A, Dzwonek A, Walker B, et al. Improving skills retention after advanced structured resuscitation training: A systematic review of randomized controlled trials. Resuscitation 2019 May;138:284-296. [doi: 10.1016/j.resuscitation.2019.03.031] [Medline: 30928503]
27. Anderson R, Sebaldt A, Lin Y, Cheng A. Optimal training frequency for acquisition and retention of high-quality CPR skills: A randomized trial. Resuscitation 2019 Feb;135:153-161. [doi: 10.1016/j.resuscitation.2018.10.033] [Medline: 30391370]

28. Thomas EJ. Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. Quality and Safety in Health Care 2004 Oct 01;13(suppl_1):i57-i64 [FREE Full text] [doi: 10.1136/qshc.2004.009811] [Medline: 15465957]
29. Bélondrade P, Lefort H, Bertho K, Perrochon J, Jost D. Tourtier J-P. Anaesth Crit Care Pain Med 2016:17 [FREE Full text] [doi: 10.1016/j.accpm.2015.06.004] [Medline: 29610057]
30. van DHP, van TL, van DHM, van DLJ. Assessment of neonatal resuscitation skills: a reliable and valid scoring system. Resuscitation 2006;71:a [FREE Full text] [doi: 10.1016/j.resuscitation.2006.04.009] [Medline: 16987590]
31. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists? Non-Technical Skills (ANTS)valuation of a behavioural marker system. Br J Anaesth 2003;90:8. [doi: 10.1037/e576952012-027] [Medline: 12697584]
32. Hagemann V, Herbstreit F, Kehren C, Chittamadathil J, Wolfertz S, Dirkmann D. Does teaching non-technical skills to medical students improve those skills and simulated patient outcome? Int J Med Educ 2017;8?13. Int J Med Educ 2017:101 [FREE Full text] [doi: 10.5116/ijme.58c1.9f0d] [Medline: 28355594]
33. Cavicchiolo M, Cavallin F, Staffler A, Pizzol D, Matediana E, Wingi O. Decision Making and Situational Awareness in Neonatal Resuscitation in Low Resource Settings. Resuscitation 2018 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.10.034] [Medline: 30391367]
34. Roh Y, Lee W, Chung H, Park Y. The effects of simulation-based resuscitation training on nurses? self-efficacy and satisfaction. Nurse Educ Today 2013;33:123 [FREE Full text] [doi: 10.1016/j.nedt.2011.11.008] [Medline: 22153054]
35. Tofil N, Peterson D, Wheeler J, Youngblood A, Zinkan J, Lara D. Repeated versus varied case selection in pediatric resident simulation. J Grad Med Educ?9 2014;6:275 [FREE Full text] [doi: 10.4300/jgme-d-13-00099.1] [Medline: 24949131]
36. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012;99:1322 [FREE Full text] [doi: 10.1002/bjs.8819] [Medline: 22961509]
37. Garden AL, Le Fevre DM, Waddington HL, Weller JM. Debriefing after simulation-based non-technical skill training in healthcare: a systematic review of effective practice. Anaesth Intensive Care 2015 May;43(3):300-308 [FREE Full text] [doi: 10.1177/0310057X1504300303] [Medline: 25943601]
38. Rondon S, Sassi F, Furquim DAC. Computer game-based and traditional learning method: a comparison regarding students? knowledge retention. BMC Med Educ 2013;13:a [FREE Full text] [doi: 10.1186/1472-6920-13-30] [Medline: 23442203]
39. Kanthan R. The impact of specially designed digital games-based learning in undergraduate pathology and medical education. Arch Pathol Lab Med 2011:135 [FREE Full text] [doi: 10.1016/j.ypat.2011.08.024] [Medline: 21204720]

Abbreviations

ANTS: Anaesthetists' Non-Technical Skills
CERAR: Comité Ethique sur la Recherche en Anesthésie Réanimation
ILCOR: International Liaison Committee on Resuscitation
NRP: Neonatal Resuscitation Program
NRPE: Neonatal Resuscitation Performance Evaluation

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by N Chaniaud, S Jung; comments to author 29.03.20; revised version received 21.04.20; accepted 14.05.20; published 11.08.20.

Please cite as:
Michelet D, Barre J, Truchot J, Piot MA, Cabon P, Tesniere A
Effect of Computer Debriefing on Acquisition and Retention of Learning After Screen-Based Simulation of Neonatal Resuscitation: Randomized Controlled Trial
JMIR Serious Games 2020;8(3):e18633
URL: http://games.jmir.org/2020/3/e18633/
doi:10.2196/18633
PMID:32780021

Original Paper

A Serious Game on the First-Aid Procedure in Choking Scenarios: Design and Evaluation Study

Imma Boada[1], PhD; Antonio Rodriguez Benitez[1], PhD; Santiago Thió-Henestrosa[1], PhD; Josep Soler[1], PhD

Graphics and Imaging Laboratory, Escola Politècnica Superior, Edifici Politècnica IV, 17003 Girona, Spain

Corresponding Author:
Imma Boada, PhD
Graphics and Imaging Laboratory
Escola Politècnica Superior, Edifici Politècnica IV
C/ de la Universitat de Girona, 6
Campus Montilivi
17003 Girona
Spain
Phone: 34 618334786
Email: imma.boada@udg.edu

Abstract

Background: Choking is one of the causes of unintentional injury death. Gaining the knowledge of the first-aid procedure that has to be applied in case of choking can increase the chances of survival of persons with choking. Serious games can be a good channel for educating people about choking scenarios and the actions to be taken to save the persons with choking.

Objective: The objective of this study is to present and evaluate the effectiveness of a serious game designed to prevent choking and to promote the first-aid procedure that needs to be applied in case of choking.

Methods: In this study, we present a serious game as a set of minigames that reproduces the main steps of the protocol for the first-aid performed in choking. In the proposed game, the player acquires the role of a helper who has to save the person in a choking emergency by applying the main steps of the protocol. Time and score restrictions are imposed to pass each minigame. To test this game, we performed a pilot study with 48 high school students. Different tests were performed to assess the students' preferences and their knowledge on choking before and after playing the proposed game. The obtained results were analyzed using Mann-Whitney *U* test when a grade variable was involved and by using Fisher exact test when 2 categorical variables were involved.

Results: The findings of our study showed that the players enjoyed the game. No statistical differences were detected when considering the gender of the player, their preferences for video games, or their previous experience in choking emergencies. By comparing the knowledge of these students before and after playing the game, we found that all the indicators of the knowledge about how to act in case of a choking emergency were improved through this serious game.

Conclusions: The findings of our study show that the proposed game is a good strategy for promoting and teaching first-aid procedures in choking emergencies to nonexperts in this field.

(JMIR Serious Games 2020;8(3):e16655) doi:10.2196/16655

KEYWORDS

choking; prevention; first-aid procedure; first-aid education

Introduction

A foreign object lodged into the throat or the windpipe may cause choking. Unless the air passage is cleared, the person with choking can lose consciousness within 3-5 minutes. In worse cases, the lack of oxygen to the brain could cause brain damage or death. In these situations, it is recommended to administer, as quickly as possible, first-aid that consists of abdominal thrusts combined with back blows [1]. Unfortunately, not everyone is aware of this procedure, and the strategies to promote the awareness of this first-aid are necessary [2].

In order to educate about the risks, prevention, and treatment of choking, different approaches have been proposed. Organizations such as the Resuscitation Council [3], the Red Cross [1], or the American Heart Association [4] have defined guidelines that describe how choking interventions should be performed both safely and effectively. These organizations also offer courses, videos, and other materials to teach the procedure of the choking rescue. In addition, many countries have initiated campaigns to educate their citizens about the prevention, risks, and treatment of choking. Special attention has been given to children since they are the most susceptible to choking [5]. The effectiveness of these initiatives has also been studied and, in most of the cases, the general conclusion is that prevention strategies are effective and contribute to decreasing the choking incidence [6-8]. Moreover, several initiatives have focused on the development of first-aid training devices. The Zoll Medical Corporation has proposed a handheld device that uses several accelerometers to monitor abdominal thrusts [9]. Recently, Watson and Zhou [10] presented BreathEZ, a smartwatch app that provides both first-aid instructions for choking and real-time tactile and visual feedback on the quality of the abdominal thrust compressions.

In this paper, we propose to tackle the problem of choking by using serious game strategies. Serious games are digital games used for purposes other than mere entertainment and are applied in different areas such as military, government, education, and health care [11]. They can recreate scenarios to experiment with situations that otherwise would be impossible in the real world owing to the required safety, cost, and time [12]. In addition, serious games enhance the development of skills such as analytical and spatial, strategic, or psychomotor [13]. Although there are a large number of serious game apps related to health care, as far as we know, there are no serious games focused only on first-aid education for choking.

Serious games have become a powerful tool to develop and acquire new knowledge and skills. Different games have been developed for health care purposes and tested in a wide range of diseases, treatments, and other related topics [14-16]. These games focus not only on expert users but also on nonexperts.

Baranowski et al [17] presented a classification of games for health in 5 categories. One category, which was centered on health care professionals, consists of games designed to provide simulation environments and virtual patients to practice and acquire relevant skills [18]. An example in this category is MyCraft, which provides virtual consultation training on tuberculosis [19] and games to practice surgeries, blood management, image-guided procedures, assessment, prevention, and treatment [20-24]. The other 4 categories, which are centered on nonexperts such as patients or general users, consist of games designed to increase knowledge, change behavior, or involve health behavior in gameplay. An example of a game that increases knowledge and changes behavior is Yummy Tricks, which is a game intended to teach healthy eating habits [25-27] or the game by Ito et al, which was developed to evaluate the dissemination of public awareness on preschool children's oral health [28,29]. Regarding games that involve behavior, exergames are the most representative of this category. These games incorporate physical activity in the gameplay and can be used for different purposes such as physical activity encouragement [30-33] or rehabilitation to recover from brain injuries, cognitive impairments [34], or motor deficiencies [35,36]. Finally, with respect to the games that influence health precursors, some examples are games that reduce stress and anxiety [37], deal with depression [38], or prepare for cancer treatments [39]. The game proposed in this paper can be classified as a game designed for nonexpert users to increase their knowledge on a health topic, particularly on first-aid survival techniques.

In the context of the first-aid techniques, different serious games have been proposed. The Virtual Heroes company [40] presented serious games for health care professionals, such as 3DiTeams [41], a first person, multiplayer training app, wherein the player is placed in a high-fidelity virtual hospital; Combat Medic [42], a web-based 3D collaborative virtual world to deal with hemorrhage, airway management, and tension pneumothorax; and HumanSim:Blast [43], wherein after a train station explosion, the player must identify and label zones on an area map, tag potential hazards, assess patient vitals, perform life-saving procedures, and triage patients. In a similar way, the BreakAway Company proposed Code Orange [44], a serious game wherein the players work in concert with the first-aid staff of a hospital to save people injured in a weapon of mass destruction event. Other games for emergency staff training are Nuclear Event Triage Challenge [45], Peninsula City [46], Burn Center [47], and CliniSpace [48]. In the field of cardiopulmonary resuscitation (CPR), Jerin et al proposed the automated external defibrillator challenge, which is a web-based serious game for teaching and training automated external defibrillation and first-aid maneuvers to lay people and emergency medical service professionals [49]. Other proposals are JUST [50], an immersive virtual reality situation training system for nonprofessional health emergency operators; MicroSim Prehospital [51] designed for prehospital training on emergency medical services; and Staying alive [52], an online 3D simulator that provides a learning experience of saving a virtual patient from cardiac arrest in 4 minutes; LISSA [53,54], which presents an emergency situation wherein CPR actions have to be applied to save the person with choking; 30:2 [55], a game designed to educate on CPR protocol to nonexperts; Relive [56], a first person 3D adventure where the player faces different rescue situations; Viva!Game [57], a web-based serious game designed to create awareness on cardiac arrest and CPR; and HeartRun [58], a mobile simulation game to train resuscitation and targeted at giving school children an understanding of this protocol. Recently, Benkhedda and Bendella [59] presented FASim, a 3D serious game that combines health care simulations with serious games and the functionalities of the multiagent systems in a single framework in order to learn the first-aid procedure for and the signs of a cardiac arrest. In the context of choking, Carvalho et al [60] proposed an Android app video game wherein different first-aid actions are presented to familiarize users with these scenarios, choking being one of them.

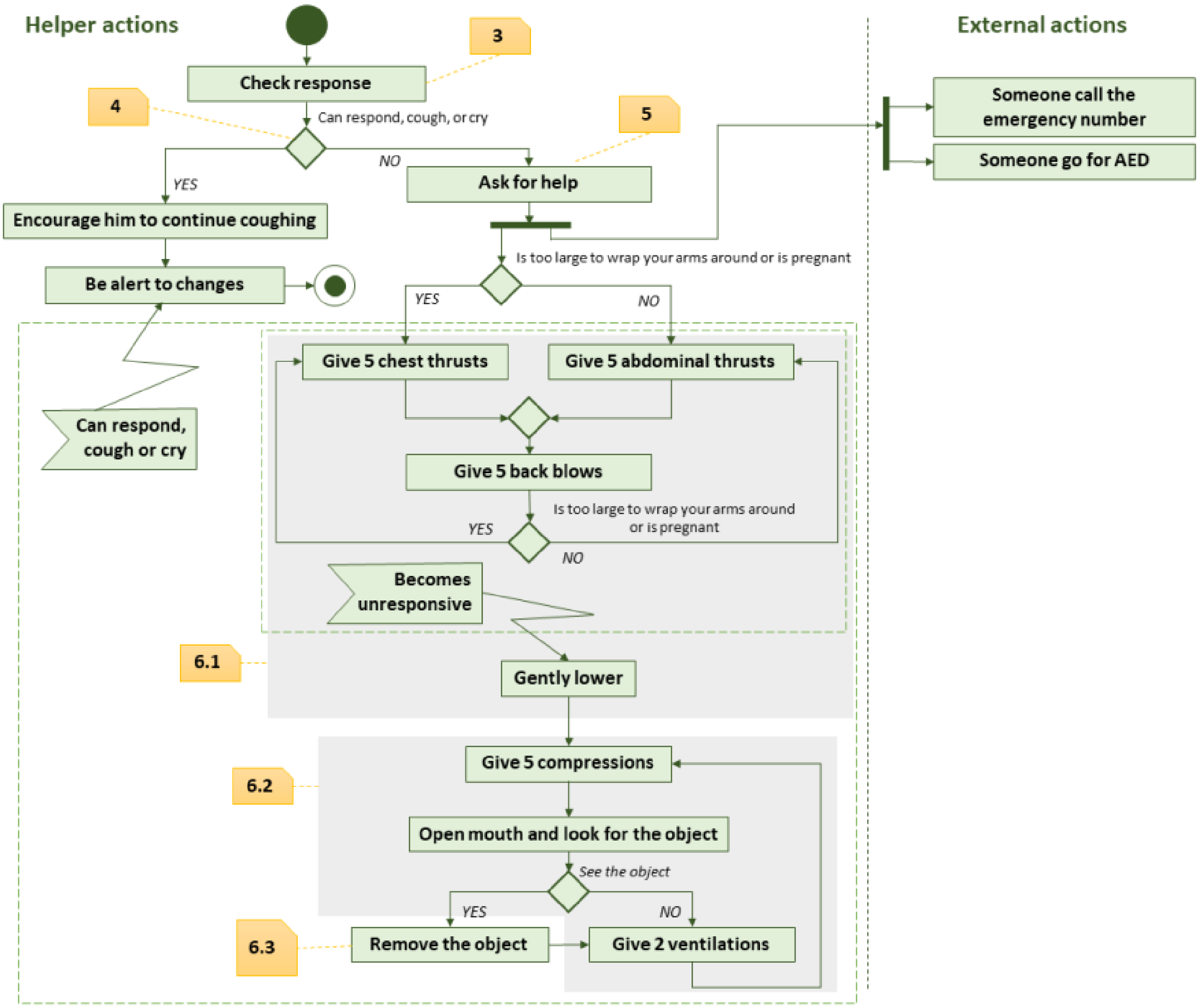
In most of the cases, the games focused on first-aid procedures have been designed for patients with health issues and the game being tailored to address their health issues adequately [61]. In general, little attention has been paid to nonexperts, although first-aid protocols would be a basic knowledge for everyone. To overcome this limitation, our aim was to exploit the advantages of serious games and use them to promote basic knowledge on choking recovery procedures. Regarding choking procedures, no games related to this topic have been proposed and this is the novel aspect of our approach. The advantages of game-based technologies over traditional education methods have been studied by many authors [62,63]. Game-based technologies are more effective because they use action instead of explanation, they are able to create personal motivation and satisfaction and accommodate multiple learning styles and skills, and they are able to provide an interactive and a decision-making context [64,65]. These facts combined with the extended use of portable gaming platforms makes computer-based games a perfect channel to promote learning contents [66] such as the ones required to educate about the choking procedure.

Methods

Main Design Decisions

In this section, we present the main considerations that have been taken into account to design the game. To create this game, the educational game development approach proposed by Torrente et al [67] was applied. This approach covers all the tasks of the game design from implementation to evaluation. It is built on 4 basic principles: (1) the procedure-centric approach that gives importance to capturing the procedural knowledge of the domain, (2) the collaboration between experts, (3) the agile development with agile tools, and (4) the low-cost game model. In our case, these principles were applied as follows. The game procedural knowledge of the domain is given by the choking protocol illustrated in Figure 1. This knowledge was supervised by the group of physicians that collaborated with our research group. The agile development was done via an iterative design and a development process that included analysis, game design, implementation, and quality assurance. Finally, instead of the 3D realistic models, our game is based on 2D animations, thereby leading to a low-cost game model.

Figure 1. Main steps of the first-aid procedure that has to be applied in a choking emergency. The numbers in the boxes indicate the relationship of the minigames 3, 4, 5, and 6 with the procedure steps. AED: automated external defibrillator.



The pedagogical approach used to design the game was based on the experiential learning theory, wherein educators aim at engaging learners in direct experience to increase their knowledge, skills, and values [68]. Experience occurs as a result of interaction between human beings and the environment in the form of thinking, seeing, feeling, handling, and doing [68]. In our case, this experience is going to take place in an artificial environment wherein a choking person has to be identified by the choking symptoms and also recovered by trying to reproduce the protocol. There are 5 instructional strategies rooted in the concept of learning through experiences. These are learning by doing, experiential, guided experiential, case-method teaching, and a combination of experiential and inquiry-based learning [64]. In our case, the proposed approach can be seen as a learning by doing strategy. The idea is to reproduce the steps of the choking protocol several times. Since just doing actions does not involve acquiring the knowledge, the game will also include instructions and feedback messages to make the actions meaningful in order to consolidate the player knowledge [69].

The last issue to be considered was how to deal with the protocol steps. From our experience in the previous games that focused on first-aid protocols [55,70], we decided to decompose the game into a set of minigames, thereby making knowledge acquisition easy for the players [71]. Therefore, the game in this study is composed of 6 minigames, the player being the first one to put in context, then presenting the elements that can cause choking, and the rest of the game is focused on the steps of the choking recovery procedure (Figure 1). The main steps of the first-aid procedure have to be applied in a choking emergency. The numbers in the boxes indicate the relationship of the minigames 3, 4, 5, and 6 with the procedure steps.

Choking Prevention Game

The 6 minigames that compose the game can be seen as submissions to reach the goal, which is to save a person from choking by applying the steps of the choking protocol. The submissions are designed to identify the choking symptoms or to apply a specific step of the protocol. In all the minigames, the player acts as a rescuer who interacts with the main character that represents the person with choking. This character appears in all the minigames in a similar scenario, with the same screen design, with time and score information on the top of the screen, as well as help messages used to highlight the relevant information of the step. To guide the player between the minigames, at the beginning of each minigame, the instructions of the step protocol are presented. There is also a help icon to access an animation that describes it. In addition, at the end of the minigame, another message communicates if the step has been achieved or not.

In all the minigames, correct actions add points and incorrect ones subtract points. The minigame is finished when the time is finished or when the maximum score is reached. Two playing modes are supported. The player can play each minigame independently to reach the maximum level or sequentially, thereby completing all the sequences of minigames with the same level of difficulty, to see the whole protocol each time. All user interactions are limited to touch and drag-and-drop actions. In this way, the players can achieve great mastery within a short period. A detailed description of each minigame is given below. For each one, first, the learning objective is presented, and then the design is proposed to achieve it. A complete demo of the game is provided on https://youtu.be/cABGCo7R2HI.

Minigame 1: Choking Prevention

In the first minigame, we focus on children who are more likely to choke than adults. The aim of this minigame is to promote 3 tips to prevent choking by using the following 3 methods: (1) avoid small and dangerous objects, (2) keep food pieces small, and (3) not move while eating. To reproduce these situations, an icon representing a child appears in the middle of the screen and different elements are placed around it. The player has to interact with these elements and carry out different actions according to the elements (Figure 2). In particular, if it appears to be a dangerous element, the player has to drag it out of the screen. If it appears to be a piece of meat, the player has to click on it 5 times to simulate cutting of small pieces. If the child icon moves from the center, the player has to drag it to the center again. This last action represents not to move while eating. In Figure 3A-3C, the different scenarios corresponding to these situations are illustrated.

Figure 2. Some of the items of the choking prevention minigame, wherein player actions vary according to the type of item.

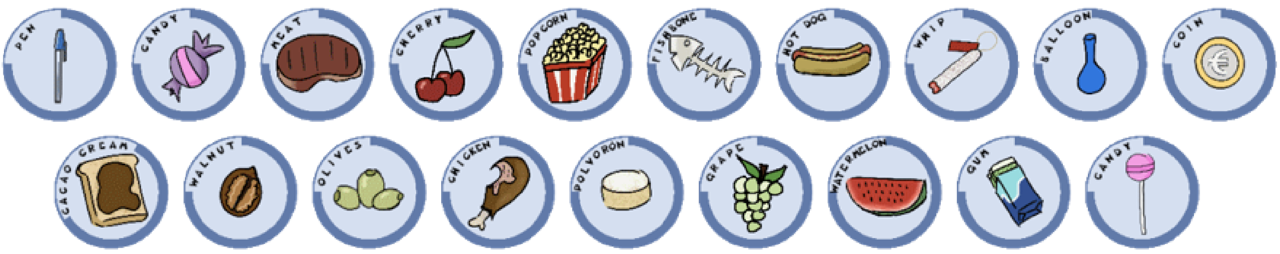
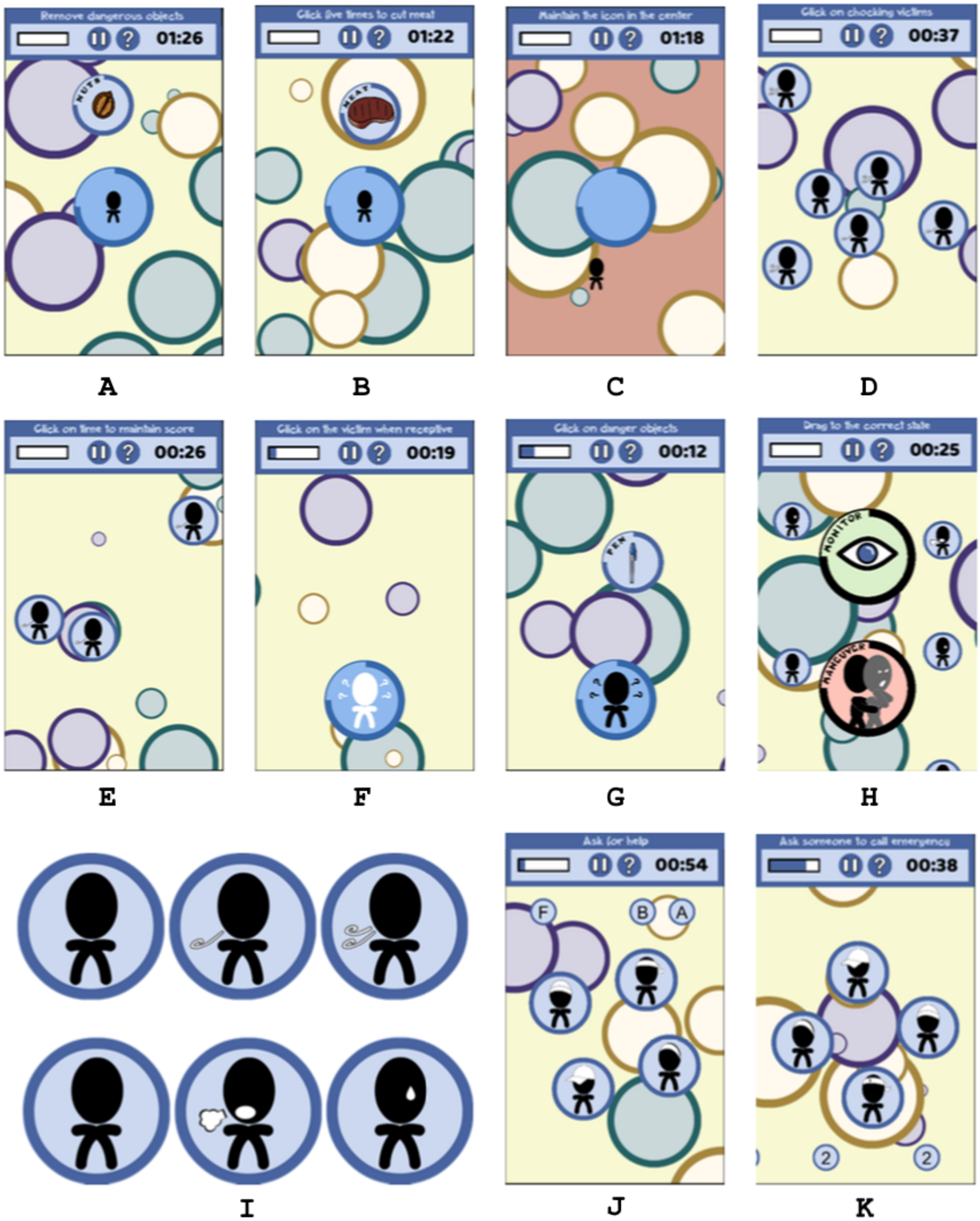


Figure 3. Choking prevention minigame. A. Avoid small and dangerous objects. B. Keep food pieces small. C. Do not move while eating. D and E. Identify choking persons minigame. F and G. Ask for response minigame. H. Identify choking symptoms minigame. I and J. Ask for help minigame. K. The icons representing breath status (no breath-breath with difficulties-normal breath) and choking symptoms (no breath-cough-cry).



Minigame 2: Identify Persons with Choking

The aim of this minigame is to identify persons with choking and remove them from the screen. On the screen appears person icons with 2, 1, or no air pathways to represent breath status as normal, with difficulty, and airway totally blocked, that is, choking person, respectively. When the icon is on the screen, its breath status can become worse (Figure 3D and Figure 3E). The player has to click on the choking persons and if these are not selected, they will disappear and the player score will decrease.

Minigame 3: Check Response

A choking person will not be able to talk but will probably communicate through signs and actions such as grabbing his or her throat. The rescuer has to know when the choking person is able to communicate or not. To reproduce this situation, in this minigame, a choking person represented as a person icon appears on the screen. The icon changes its color from black to white to indicate that the choking person is receptive to be asked or not, respectively. In addition, danger elements going to the choking person appear on the screen. The player has to drag these objects out from the screen and click on the choking person when the icon is receptive be asked. If a dangerous object arrives to the choking person, the icon cannot be asked and the player score will decrease. The score also decreases when no receptive choking person is asked. In Figure 3F and Figure 3G, screenshots of this minigame are presented.

Minigame 4. Identify Choking Symptoms

A choking person typically has a panicked, confused, or surprised facial expression and usually place hands on the throat. If the airway is not totally blocked, choking persons will be able to cough or make squeaking noises while trying to breathe. If the airway is totally blocked, the choking persons will not be able to speak, cry, or cough, and their skin color will range from red to pale owing to the lack of oxygen. In this minigame, person icons appear on the screen and some of them represent choking symptoms. The player has to identify icons and separate choking persons from the others. Points are lost in case of incorrect or nonclassification of choking persons. A screenshot of this minigame is presented in Figure 3H, and the icons representing the different symptoms are shown in Figure 3K.

Minigame 5. Ask for Help

After the identification of the person with choking and confirming the person with choking, the player has to ask for help and call an emergency number.

In this minigame, person icons as well as letters on the top and numbers on the bottom will appear on the screen. The player has to select one of the icons and write *help* by combining selected letters from the top. The icons and letters are continuously moving. The same procedure has to be done to call emergency numbers by selecting one of the icons, not necessarily the same, and attaching the correct emergency numbers. The player ends the game when *help* and the correct emergency number are written. The game screens are shown in Figure 3I and Figure 3J.

Minigame 6. Choking Maneuver

This minigame represents the most important part of the choking rescue protocol. To reproduce it, we have divided this minigame into 3 parts that recreate back blows and abdominal thrust, CPR, and object removal. To pass this minigame, the player has to pass the 3 parts. The description of each part is presented below.

Perform Back Blows and Abdominal Thrusts

For adults and children with choking, to force the object out of the airway, the helper has to give a combination of 5 back blows between the shoulder blades followed by 5 abdominal inward and upward thrusts just above the navel. To reproduce this situation, 2 silhouettes representing the choking person and the helper and icons representing different hand positions will appear on the screen (Figure 4). First, the player has to put the choking person in the correct position by clicking on the head and dragging to the right side (Figure 5A) and then select the correct hand position represented in one of the different icons that will appear (Figure 5B). Then, the player has to give 5 clicks on the correct position of the choking person (Figure 5C) to simulate the 5 back blows. The player has to place the choking person in the initial position by clicking on the head and dragging to the left side (Figure 5D). The player has to perform 5 abdominal thrusts by selecting the correct position of the hands (Figure 5B) and then by clicking 5 times in the correct position (Figure 5E). This process is repeated until the choking person ejects the object or passes out. If the person ejects the object, a new choking person will appear. If the choking person passes out, the second part of the minigame, that is, CPR, will start.

Figure 4. Icons representing different hand positions.

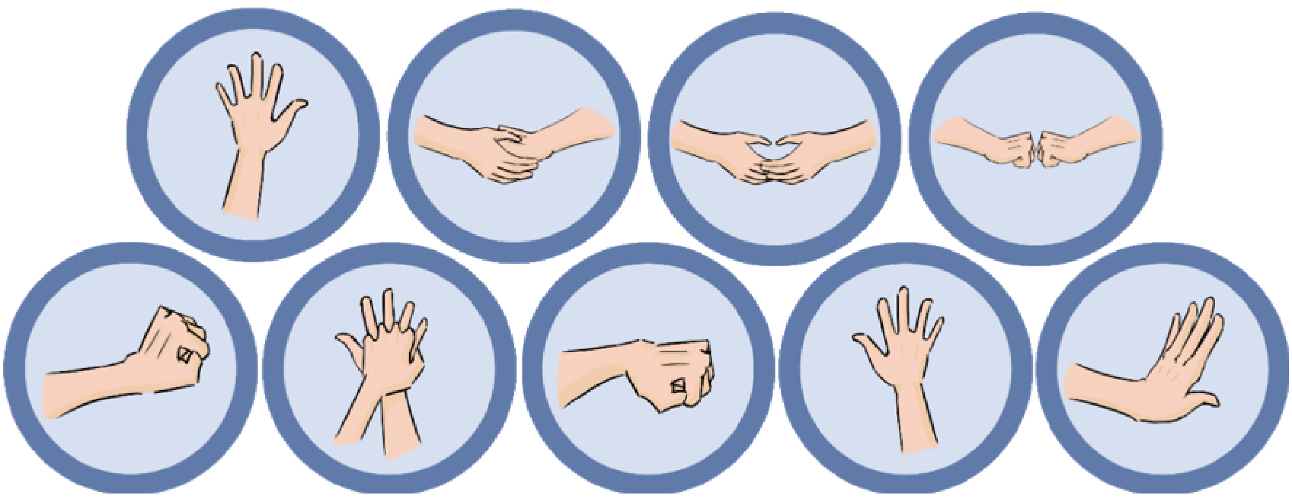
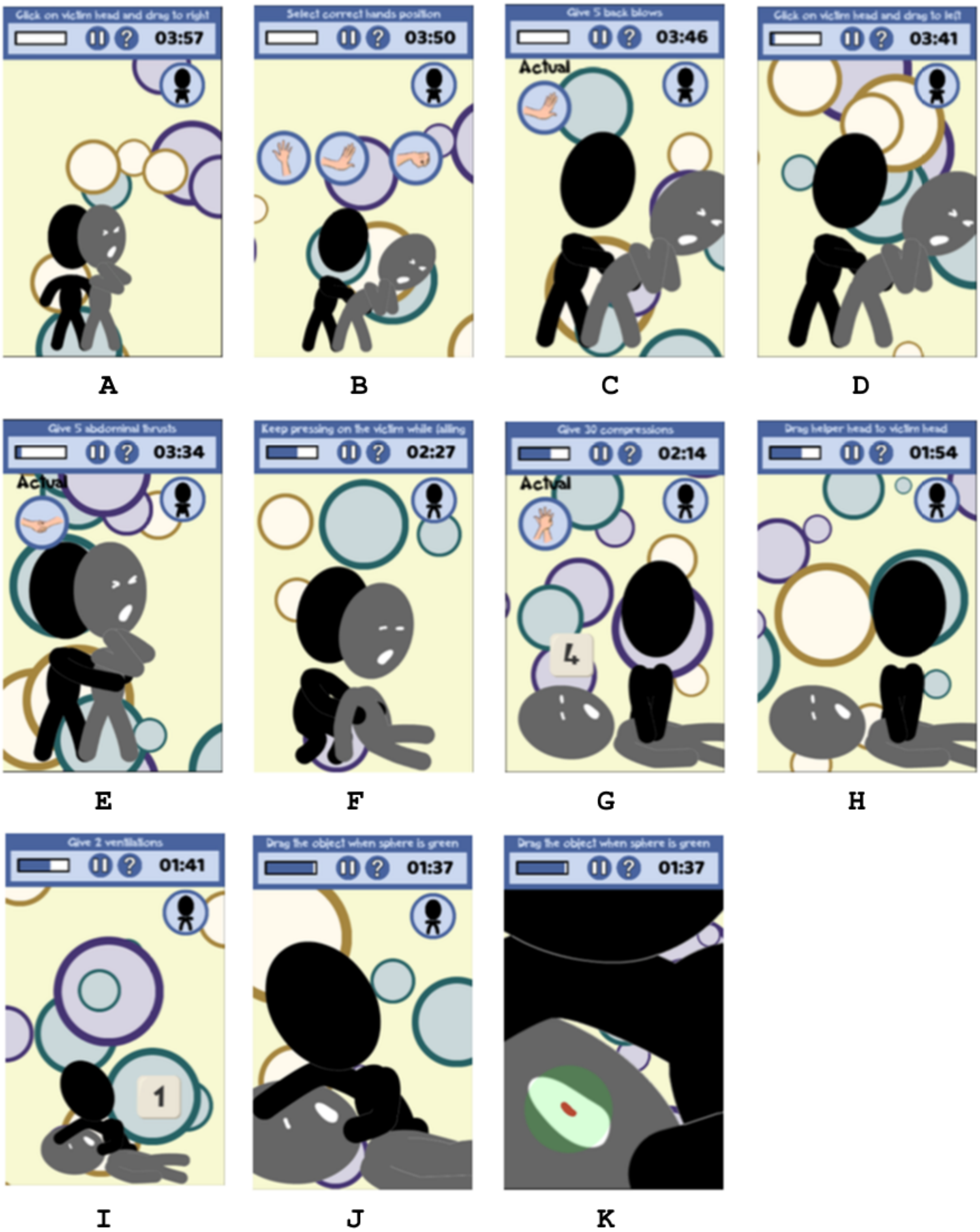


Figure 5. Choking maneuver screens corresponding to the different parts of the minigame. A-E. Back blows and abdominal thrusts. B, F, G, H, J. Cardiopulmonary resuscitation. I. Remove the object. K. Rescue breaths.



Perform CPR

When the choking person is unresponsive because back blows and abdominal thrusts have been unsuccessful, the player has to perform the CPR, which combines 30 chest compressions with 2 rescue breaths. First of all, the player has to gently lower the choking person by touching the body (Figure 5F). At this point of the minigame, the choking person and the helper stay in the position of the CPR protocol. To start, the player has to select the icon representing the correct hand position to perform chest compressions. Then, the player has to click 30 times on the chest of the choking person following the correct rhythm (Figure 5G). After each set of chest compressions and before rescue breaths, the player has to click on the mouth of the choking person and look for the object. If the player sees the object, the third part of the minigame starts (Figure 5H). If the object is not visible, the player has to perform 2 rescue breaths by touching the mouth of the choking person following the correct rhythm (Figure 5K). Again, a new sequence of compressions has to be performed and the process is repeated until help arrives.

Remove the Object

If the helper sees the object in the mouth of the of the choking person, it has to be removed. To reproduce this situation, the object appears in the mouth of the choking person with a sphere behind it. The sphere changes its color and the player can only move the object when it is green (Figure 5I and Figure 5J). If the player moves the object when the sphere has another color, the object will fall. The object has to be removed in the given time.

Designed Study and Statistical Analysis

To test the game, a sample population consisting of 48 high school students from a summer camp of our university was considered for this study. Our laboratory, the Graphics and Imaging Laboratory, participated in this camp for a 2-h session that was carried out in a computer laboratory. The aim of the session was to introduce high school students to video games and serious games. After the first introduction of our research, we asked the students to answer the questionnaire presented in Textbox 1. Then the students had 30 minutes to play the proposed game as an example of a serious game. Each student had a computer to play, and no introduction to the game was given; the students were just asked to play the game. During the session, we observed them so that we could detect the difficulties that the students encountered in the game. After the session, we asked them to answer the new questionnaire presented in Textbox 2.

To detect significant statistical differences, we used the Mann-Whitney *U* test when a grade variable was involved and we used the Fisher exact test when 2 categorical variables were involved.

Textbox 1. Test 1 questionnaire before playing the game.

Questionnaire to be completed before the game

1. Select gender: Male/female

2. Age: Insert the number

3. Do you like video games? Yes/No

4. Do you know what is choking? Yes/No

5. Grade your knowledge of the choking first-aid protocol from 1 to 5 (1 being the least and 5 being the highest):

6. Have you ever been in a choking emergency? Yes/No

7. In case of choking, to remove the throat object you have to (select only 1 correct option):

- Perform abdominal thrusts until object expulsion
- Perform back blows until object expulsion
- Repeat 5 back blows and 5 abdominal thrusts until object expulsion
- Repeat 10 back blows and 10 abdominal thrusts until object expulsion.

8. In which of the following cases of choking, would the cardiopulmonary resuscitation (CPR) protocol be applied? (select only 1 correct option):

- In all the cases, as it is a part of the choking first-aid protocol
- In case of unconscious choking persons
- Never, as the object causing choking does not need to be removed by CPR.

9. Write the emergency number.

Textbox 2. Test 2 questionnaire after playing the game.

Questionnaire to be completed after the game

1. Do you know what is choking? Yes/No

2. Grade your knowledge of the choking first-aid protocol from 1 to 5 (1 being the least and 5 being the highest):

3. In case of choking, to remove the throat object you have to (select only 1 correct option):

- Perform abdominal thrusts until object expulsion
- Perform back blows until object expulsion
- Repeat 5 back blows and 5 abdominal thrusts until object expulsion
- Repeat 10 back blows and 10 abdominal thrusts until object expulsion.

4. In which of the following cases of choking, would the cardiopulmonary resuscitation (CPR) protocol be applied? (select only 1 correct option):

- In all the cases, as it is a part of the choking first-aid protocol
- In case of unconscious choking persons
- Never, as the object causing choking does not need to be removed by CPR.

5. Write the emergency number.

6. Grade your experience of the game from 1 (completely disagree) to 5 (completely agree) for each of the following categories:

- The game is entertaining.
- I like the game.
- The complexity of the game is correct.
- The help indications provided by the game are enough.
- The game describes how to proceed in a choking emergency.
- After playing the game, I know how to proceed in a choking emergency.

Results

Categories of the Findings

The obtained results are presented in 3 parts. The first part describes the sample population, the second one describes the results obtained from the game performance evaluation, and the third one compares the results obtained before and after playing the game.

Test Population

Our test population of 48 high school students consisted of 16 males (33%) and 32 females (67%). Their mean age was 15.42 (SD 0.74) years. Of these 48 students, 29 liked video games (60%) while 19 did not like video games (40%), and12 encountered a choking emergency (25%) while 36 never encountered this emergency situation (75%).

Evaluation of the Game Performance

Firstly, we present the results obtained from the answers related to the game such as the complexity and the degree of entertainment (Test 2 question 6). Then, we describe the results considering the gender of the player, video game preferences, and choking emergency experience.

In Table 1, the descriptive statistics of the answers to the questions related to the game are shown. Although, the answers are not extremely positive, we can observe that they tend to be more positive than negative. Note that all the medians are above 3; the median of *the help indications provided in the game are enough* (3.8) was the highest followed by the medians of *the game describes how to proceed in a choking emergency* (3.58), *the game is entertaining* (3.5), and *the complexity of the game is correct* (3.46). These results show that the primary aim of the game, that is teaching the procedure that has to be applied in case of choking, was achieved. In addition, we saw that the players enjoyed the game; it was considered to be not boring and not difficult. Our results also show that the game effectively described how to proceed in a choking emergency.

In Table 2, the descriptive statistics of the answers to the questions related to the game based on the player gender are shown. No statistically significant differences were found except with regard to the complexity of the game. All the items showed a small effect size, except the complexity of the game, which showed a moderate effect size. Regarding the complexity of the game, more males than females agreed that that complexity was appropriate. The results based on whether the students liked video games or not are presented in Table 3, and the results based on whether the students had a previous experience of choking emergency are presented in Table 4. No statistically significant differences were found in any of the cases, and the effect size was small. Therefore, we can consider that this serious game on first-aid for choking fits the different player profiles.

Table 1. Descriptive statistics of the answers to the questions related to the game.

Answers to the questions (grades 1-5)	Median (Q1, Q3)
The game is entertaining	3.5 (3, 4)
I like the game	3.02 (2, 4)
The complexity of the game is correct	3.46 (3, 4)
The help indications provided by the game are enough	3.8 (2, 4)
The game describes how to proceed in a choking emergency	3.58 (2.75, 5)
After playing the game, I know how to proceed in a choking emergency	3.13 (2, 4)

Table 2. Descriptive statistics of the answers to the questions related to the game (grades 1-5) by each gender.

Answers, Gender	Median (Q1, Q3)	*P* value	Effect size
The game is entertaining		.21	0.182
Females[a]	4 (3, 4)		
Males[b]	3 (2, 4)		
I like the game		>.99	0.0016
Females	3 (2, 4)		
Males	3 (2, 4)		
The complexity of the game is correct		.03	0.324
Females	3 (3, 4)		
Males	4 (3.75, 4.25)		
The help indications provided by the game are enough		.85	0.293
Females	3 (2, 4)		
Males	3 (2, 4)		
The game describes how to proceed in a choking emergency		.69	0.059
Females	4 (2.75, 4.25)		
Males	4 (3, 5)		
After playing the game, I know how to proceed in a choking emergency		.69	0.060
Females	3 (2, 4)		
Males	3 (3, 4)		

[a]Number of females in the study=32.

[b]Number of males in the study=16.

Table 3. Descriptive statistics of the answers to the questions related to the game (grades 1-5) depending on whether the students like video games or not.

Answers, Video game preference category	Median (Q1, Q3)	*P* value	Effect size
The game is entertaining		.89	0.021
Not like[a]	3 (3, 4)		
Like[b]	4 (3, 4)		
I like the game		.17	0.198
Not like	3 (2, 3)		
Like	3 (2, 4)		
The complexity of the game is correct		.19	0.192
Not like	3 (3, 4)		
Like	4 (3, 4)		
The help indications provided by the game are enough		.64	0.069
Not like	3 (2, 4)		
Like	3 (2, 4)		
The game describes how to proceed in a choking emergency		.64	0.069
Not like	4 (3, 4.5)		
Like	3 (3, 5)		
After playing the game, I know how to proceed in a choking emergency		.36	0.133
Not like	3 (2, 4)		
Like	3 (3, 4)		

[a]Number of students who did not like playing video games=19.

[b]Number of students who liked playing video games=29.

Table 4. Descriptive statistics of the answers to the questions related to the game (grades 1-5) depending on whether the students had previously experienced a choking emergency.

Answers, Experience of a choking emergency	Median (Q1, Q3)	*P* value	Effect size
The game is entertaining		.68	0.061
No[a]	4 (3, 4)		
Yes[b]	3.5 (2.75, 4)		
I like the game		.87	0.025
No	3 (2, 3)		
Yes	3 (2.75, 4)		
The complexity of the game is correct		.57	0.084
No	4 (3, 4)		
Yes	3 (3, 4)		
The help indications provided by the game are enough		.81	0.036
No	3 (2, 4)		
Yes	3 (2, 4)		
The game describes how to proceed in a choking emergency		.49	0.100
No	4 (3, 5)		
Yes	3.5 (2.75, 4.25)		
After playing the game, I know how to proceed in a choking emergency		.72	0.053
No	3 (2, 4)		
Yes	3 (2.7, 4)		

[a]Number of students who had never been in a choking emergency=36.

[b]Number of students who been in a choking emergency=12.

Protocol Knowledge Before and After Playing

Focusing on the main steps of the choking protocol (Textbox 1 questions from 4 to 9), we compared the answers obtained before and after playing the game.

In Table 5, the self-impression of the knowledge of the choking protocol before and after playing the game is shown. These results are presented for the complete sample and by gender, game preference, and experience of choking scenarios. The findings in all the analyses showed an improvement after playing the game, with small effect sizes, except in the groups of males and those who liked video games, which showed moderate effect size, and in those with experience in choking scenarios, which is large.

In addition, in Table 6 and Table 7, the results of *how the students have to act in case of choking to remove the object in the throat* and *when the resuscitation protocol has to be applied* are shown. In these 2 analyses, we compared the correct answers with the incorrect answers. The results are presented for the whole sample and by gender, game preference, and experience of choking scenarios. All the indicators of the knowledge about how to act in case of a choking emergency improved after playing the game. When the same questions were analyzed by gender, game preference, and experience of the choking scenarios, all the groups achieved a significant improvement, except in the action of *removing the object from the throat*. In this case, males and those who had a previous experience of choking scenarios did not show significant improvement. In the situation of *when the resuscitation protocol has to be applied in a choking emergency*, males did not show significant improvement.

Table 5. Self-impression of the knowledge of the choking protocol before and after playing the game (grades from 1 to 5).

Student categories, Time of answering the questionnaire	Median (Q1, Q3)	*P* value	Effect size
Complete sample population, N=48		<.001	0.120
Before playing	3 (2, 4)		
After playing	4 (3, 5)		
Females, n=32		.002	0.217
Before playing	3 (2.75, 3.25)		
After playing	3 (3, 4)		
Males, n=16		.002	0.345
Before playing	3 (2, 4)		
After playing	4 (3.75, 5)		
Students who did not like playing video games, n=19		.03	0.136
Before playing	3 (2.5, 3)		
After playing	3 (3, 3.5)		
Students who liked playing video games, n=29		<.001	0.407
Before playing	3 (2, 4)		
After playing	4 (4, 5)		
Students who did not have a previous experience of a choking emergency, n=36		<.001	0.147
Before playing	3 (2, 3.25)		
After playing	4 (3, 4)		
Students who had a previous experience of a choking emergency, n=12		.047	0.707
Before playing	3 (3, 4.25)		
After playing	4.5 (3, 5)		

Table 6. Statistical analysis of the responses of the action in case of choking before and after playing the game.

Student category and subcategories, Responses before the game (n)	Responses after the game (n)		*P* value
	Correct	Incorrect	
Complete sample, N=48			<.001
Correct	28	0	
Incorrect	12	8	
Gender			
Females, n=32			<.001
Correct	19	0	
Incorrect	6	7	
Males, n=16			.44
Correct	9	0	
Incorrect	6	1	
Video game preference			
Do not like, n=19			.02
Correct	11	0	
Incorrect	4	4	
Like, n=29			.02
Correct	17	0	
Incorrect	8	4	
Previous experience in a choking emergency			
No, n=36			.003
Correct	21	0	
Incorrect	9	6	
Yes, n=12			.15
Correct	7	0	
Incorrect	3	2	

Table 7. Statistical analysis of the responses regarding the resuscitation protocol before and after playing the game.

Student category and subcategories, Responses before the game (n)	Responses after the game (n)		*P* value
	Correct	Incorrect	
Complete sample, N=48			<.001
Correct	27	1	
Incorrect	8	12	
Gender			
Females, n=32			<.001
Correct	20	1	
Incorrect	2	9	
Males, n=16			.21
Correct	7	0	
Incorrect	6	3	
Video game preference			
Do not like, n=19			.009
Correct	11	1	
Incorrect	2	5	
Like, n=29			.001
Correct	16	0	
Incorrect	6	7	
Previous experience in a choking emergency			
No, n=36			.003
Correct	21	1	
Incorrect	7	7	
Yes, n=12			.015
Correct	6	0	
Incorrect	1	5	

Discussion

Serious games have become a useful training technology in the health care profession and for patients to learn about the procedures involved in health care. Serious games are applicable in different fields such as surgery, odontology, cardiology, nursing, diabetes, psychology, or first-aid [16,72-74]. First-aid, triage, and mass emergency are the most popular fields, wherein games have been developed for training residents, medical doctors, or students. However, little attention has been paid to general players with no health issues. In the context of first-aid education, the majority of the proposed games have focused on the CPR protocol [53-55]. However, no games have focused on the procedure to be applied in case of choking. To overcome this limitation, in this paper, we have proposed a game to introduce the main steps of the first-aid procedure in cases of choking for the general population. This game was tested on a sample population of 48 high school students. Although tests have been done with young people, this game in our study has been designed for the general public with no age restrictions.

In the proposed game, the player acquires the role of a helper who has to save a person in a choking emergency by detecting the choking symptoms and applying the main steps of the protocol. This serious game has been designed as a set of minigames, with each minigame focused on a single concept of the protocol. Several authors have demonstrated that the use of serious minigames requires less time to master the game, which has a positive impact on the learning process. In addition, minigames make the study of a subject from different angles more encouraging [75-77]. These advantages have been exploited in this game in our study. Our minigames have very simple mechanics and the interaction is reduced to touch or drag-and-drop actions. The simplicity of this game allows the player to achieve great mastery in a short period of time. The actions reproduce some steps of the protocol that will be transmitted by applying the learning-by-doing strategy. To relate the performed actions with the protocol step, different messages are also provided to the player in the form of instructions or feedback. In addition, there is a help icon to access an animation that presents the gameplay mechanics. Time and score restrictions are imposed to pass each minigame. These restrictions increase when the game advances with the aim of creating an addiction [76].

In the tests, we observed that the players do not access the provided help icon. The players preferred a trial-and-error strategy to discover the game mechanics than the provided animations. However, in some cases, the player reached the end of the game without performing any correct action. It seems that the players prefer the trial-and-error strategy rather than the help option. To avoid this situation, we modified the help activation in such a way that if the game detects a period of time with no correct actions, the help animation is automatically activated. The idea is to reduce this trial-and-error period in order to be more effective with the minigame. In addition, if the help option is on (playing with help mode) when a new game starts, the help animation is automatically activated with no game interruption.

In the tests, we also evaluated player impressions about the game complexity, enjoyability, etc. From the results, we observed that players enjoyed the game irrespective of their gender, their preferences for video games, or their previous experience in a choking emergency. Regarding the acquired knowledge, to evaluate it, pretests and posttests were carried out. We observed that the knowledge on choking and the first-aid procedure for choking was improved through the game. Therefore, the idea of using minigames to introduce the choking concepts becomes ideal for creating an awareness of the topic in an engaging and a quick manner [78]. From these results, we considered that by focusing on protocols composed of different steps, this same strategy can be applied by simply designing serious minigames for each step and integrating all of them in a common story or in a common scenario. In this way, the whole objective of learning a protocol can be decomposed into a set of subobjectives. Although the obtained results were satisfactory, we consider that different improvements need to be done. In our study, we focused on high school students, which is a limiting factor since all of them are in the same range of ages. To overcome this limitation, our idea is to extend the study to a more general population. In addition, we want to evaluate the minigames independently. We detected some game preferences and some difficulties in some minigames but these were obtained from the visual observation and we consider that these observations are not robust enough to be included in the paper. We aim to design a new experiment to carry out this evaluation. Moreover, we have not considered any particular player's capabilities, needs, and interests. We also want to consider players with visual impairments [70].

In conclusion, serious games are increasingly gaining attention in health care to complement and promote the training of experts in the field. However, little attention has been paid to the use of serious games as a tool to promote health in a nonexpert population. Our study proposes a serious game with the aim of educating on the first-aid procedure for choking. Our game introduces the main steps of the procedure as a set of minigames. It has been tested in a pilot study, and very promising results have been obtained. The students enjoyed the game and, more importantly, their knowledge on the first-aid for choking was found to be greatly improved. We conclude that serious games are a good strategy to promote first-aid knowledge to nonexperts.

Acknowledgments

This work was supported by the Catalan Government (Grant No. 2017-SGR-1101) and by the Spanish Government (Grant No. PID2019-106426RB-C31).

Authors' Contributions

ARB designed and developed the game and conducted all the experiments. IB designed the game and the experiments and drafted the paper. STH designed the experiments and conducted all statistical analyses. JS drafted the paper. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

References

1. The British Red Cross Society. British Red Cross First Aid Resources. URL: https://www.redcross.org.uk/first-aid/ [accessed 2020-05-20]
2. Nichols BG, Visotcky A, Aberger M, Braun NM, Shah R, Tarima S, et al. Pediatric exposure to choking hazards is associated with parental knowledge of choking hazards. Int J Pediatr Otorhinolaryngol 2012 Feb;76(2):169-173. [doi: 10.1016/j.ijporl.2011.10.018] [Medline: 22169464]
3. Resuscitation Council (UK). Lifesaver learning. URL: https://www.resus.org.uk/apps/lifesaver/ [accessed 2020-05-25]
4. American Heart Association. AHA information. 2020 May 25. URL: https://www.heart.org/ [accessed 2020-05-25]
5. Thamboo A, Nguyen T, Ludemann JP. Choking prevention: shortcomings of traditional public education campaigns, prompting the development of web-based interactive teaching tools for preteens, teenagers, and adults. J Otolaryngol Head Neck Surg 2009 Dec;38(6):646-651. [Medline: 19958728]
6. Bentivegna KC, Borrup KT, Clough ME, Schoem SR. Basic choking education to improve parental knowledge. Int J Pediatr Otorhinolaryngol 2018 Oct;113:234-239. [doi: 10.1016/j.ijporl.2018.08.002] [Medline: 30173993]

7. Karatzanis AD, Vardouniotis A, Moschandreas J, Prokopakis EP, Michailidou E, Papadakis C, et al. The risk of foreign body aspiration in children can be reduced with proper education of the general population. Int J Pediatr Otorhinolaryngol 2007 Feb;71(2):311-315. [doi: 10.1016/j.ijporl.2006.10.020] [Medline: 17161875]
8. Dowd MD. Choking in Children: What to Do and How to Prevent. Pediatr Ann 2019 Sep 01;48(9):e338-e340. [doi: 10.3928/19382359-20190819-01] [Medline: 31505005]
9. Totman MH. Method of measuring abdominal thrusts for clinical use and training. US Patent 9,183,762 B2. Nov.: ZOLL Medical Corporation; 2015. URL: https://patents.google.com/patent/US9183762 [accessed 2020-05-20]
10. Watson A, Zhou G. BreathEZ: Using smartwatches to improve choking first aid. Smart Health 2019 Aug;13:100058. [doi: 10.1016/j.smhl.2018.07.026]
11. Michael D, Chen S. Serious Games: Games That Educate, Train, and Inform. US: Thomson Course Technology; 2006.
12. Squire K, Jenkins H. Harnessing the power of games in education. Sematic Scholar. 2003. URL: https://www.semanticscholar.org/paper/HARNESSING-THE-POWER-OF-GAMES-IN-EDUCATION-Squire-Jenkins/718e12db4a77643b1584419eaa8983c655cbf5f9 [accessed 2019-08-03]
13. Mitchell A, Savill-Smith C. The Use of Computer and Video Games for Learning: A Review of the Literature. London: Learning and Skills Development Agency; 2004. URL: https://dera.ioe.ac.uk/5270/7/041529_Redacted.pdf [accessed 2019-08-02]
14. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
15. Laamarti F, Eid M, El Saddik A. An Overview of Serious Games. International Journal of Computer Games Technology 2014;2014:1-15. [doi: 10.1155/2014/358152]
16. Ricciardi F, De Paolis LT. A Comprehensive Review of Serious Games in Health Professions. International Journal of Computer Games Technology 2014;2014:1-11. [doi: 10.1155/2014/787968]
17. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for Health for Children-Current Status and Needed Research. Games Health J 2016 Feb;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
18. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
19. Applied Research Associates. MYCRAFT. Virtual Heroes website. URL: https://www.virtualheroes.com/portfolio/Medical/Mycraft [accessed 2020-05-25]
20. Qin J, Chui Y, Pang W, Choi K, Heng P. Learning blood management in orthopedic surgery through gameplay. IEEE Comput Graph Appl 2010;30(2):45-57. [doi: 10.1109/MCG.2009.83] [Medline: 20650710]
21. Sabri H, Cowan B, Kapralos B, Moussa F, Cristanchoi S, Dubrowski A. Off-pump coronary artery bypass surgery procedure training meets serious games. 2010 Presented at: 9th IEEE International Symposium on Haptic Audio Visual Environments and Games; 2010; Phoenix p. 1-5. [doi: 10.1109/have.2010.5623995]
22. Chan W, Qin J, Chui Y, Heng P. A serious game for learning ultrasound-guided needle placement skills. IEEE Trans Inf Technol Biomed 2012 Nov;16(6):1032-1042. [doi: 10.1109/TITB.2012.2204406] [Medline: 22711779]
23. Applied Research Associates. HumanSim Critical Decisions: A Skins Game. Virtual Heroes website. URL: https://virtualheroes.com/portfolio/Medical/HumanSim-Critical-Decisions--A-Skins-Game [accessed 2020-05-20]
24. University of South Wales. Learn to ECG. Virtual ECG. URL: http://ecg.southwales.ac.uk/ [accessed 2020-05-25]
25. Papastergiou M. Exploring the potential of computer and video games for health and physical education: A literature review. Computers & Education 2009 Nov;53(3):603-622. [doi: 10.1016/j.compedu.2009.04.001]
26. Makhlysheva A, Arsand E, Hartvigsen G. Review of Serious Games for People with Diabetes. In: Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice. US: IGI Globall; 2016:412-447.
27. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
28. Ito C, Marinho Filho AV, Ito M, Azevedo MM, de Almeida MA. Preliminary evaluation of a serious game for the dissemination and public awareness on preschool children's oral health. Stud Health Technol Inform 2013;192:1034. [Medline: 23920808]
29. Inglés-Camats G, Presno-Rivas MM, Antonijoan M, Garcia-Panella O, Forrest T. Yummy tricks: a serious game for learning healthy eating habits. Stud Health Technol Inform 2012;172:185-190. [Medline: 22910522]
30. Cacciata M, Stromberg A, Lee J, Sorkin D, Lombardo D, Clancy S, et al. Effect of exergaming on health-related quality of life in older adults: A systematic review. Int J Nurs Stud 2019 May;93:30-40. [doi: 10.1016/j.ijnurstu.2019.01.010] [Medline: 30861452]
31. Staiano AE, Calvert SL. The promise of exergames as tools to measure physical health. Entertain Comput 2011 Jan 01;2(1):17-21 [FREE Full text] [doi: 10.1016/j.entcom.2011.03.008] [Medline: 23378860]
32. Vernadakis N, Papastergiou M, Zetou E, Antoniou P. The impact of an exergame-based intervention on children's fundamental motor skills. Computers & Education 2015 Apr;83:90-102. [doi: 10.1016/j.compedu.2015.01.001]

33. Ruggiero L, Seltzer ED, Dufelmeier D, McGee Montoya A, Chebli P. MyPlate Picks: Development and Initial Evaluation of Feasibility, Acceptability, and Impact of an Educational Exergame to Help Promote Healthy Eating and Physical Activity in Children. Games Health J 2020 Jun;9(3):197-207. [doi: 10.1089/g4h.2019.0056] [Medline: 32293906]
34. van der Kuil MNA, Visser-Meily JMA, Evers AWM, van der Ham IJM. A Usability Study of a Serious Game in Cognitive Rehabilitation: A Compensatory Navigation Training in Acquired Brain Injury Patients. Front Psychol 2018;9:846 [FREE Full text] [doi: 10.3389/fpsyg.2018.00846] [Medline: 29922196]
35. Preston N, Weightman A, Gallagher J, Levesley M, Mon-Williams M, Clarke M, et al. A pilot single-blind multicentre randomized controlled trial to evaluate the potential benefits of computer-assisted arm rehabilitation gaming technology on the arm function of children with spastic cerebral palsy. Clin Rehabil 2016 Oct;30(10):1004-1015. [doi: 10.1177/0269215515604699] [Medline: 26370148]
36. Rego P, Moreira P, Reis L. Serious games for rehabilitation: A survey and a classification towards a taxonomy. In: Serious games for rehabilitation: A survey and a classification towards a taxonomy. Inth Iberian Conference on Information Systems and Technologies (CISTI). Santiago de Compostela, Spain: IEEE; 2010 Presented at: 5th Iberian Conference on Information Systems and Technologies (CISTI); june 2010; Santiago de Compostela p. 1-6 URL: https://ieeexplore.ieee.org/document/5556674
37. Carlier S, Van der Paelt S, Ongenae F, De Backere F, De Turck F. Empowering Children with ASD and Their Parents: Design of a Serious Game for Anxiety and Stress Reduction. Sensors (Basel) 2020 Feb 11;20(4) [FREE Full text] [doi: 10.3390/s20040966] [Medline: 32054025]
38. Dias LPS, Barbosa JLV, Vianna HD. Gamification and serious games in depression care: A systematic mapping study. Telematics and Informatics 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
39. Govender M, Bowen RC, German ML, Bulaj G, Bruggers CS. Clinical and Neurobiological Perspectives of Empowering Pediatric Cancer Patients Using Videogames. Games Health J 2015 Oct;4(5):362-374 [FREE Full text] [doi: 10.1089/g4h.2015.0014] [Medline: 26287927]
40. Applied Research Associates. Free Education and Training Downloads For Real Heroes. Virtual Heroes. URL: https://www.virtualheroes.com/ [accessed 2020-05-20]
41. Applied Research Associates. Virtual Heroes. 3DiTeams. 2007. URL: https://www.virtualheroes.com/portfolio/Medical/3DiTeams [accessed 2020-05-25]
42. Applied Research Associates. Combat Medic. Virtual Heroes. 2014. URL: https://www.virtualheroes.com/content/combat-medic [accessed 2020-05-20]
43. Applied Research Associates B. HumanSim Blast. Virtual Heroes. 2011. URL: https://www.virtualheroes.com/portfolio/Medical/HumanSim--Blast [accessed 2020-05-20]
44. Code Orange. BreakAway Games. 2016. URL: https://www.breakawaygames.com/games/code-orange/ [accessed 2020-05-25]
45. Bergeron BP. Learning & retention in adaptive serious games. Stud Health Technol Inform 2008;132:26-30. [Medline: 18391250]
46. LeRoy Heinrichs W, Youngblood P, Harter PM, Dev P. Simulation for team training and assessment: case studies of online training with virtual worlds. World J Surg 2008 Feb;32(2):161-170. [doi: 10.1007/s00268-007-9354-2] [Medline: 18188640]
47. Kurenov SN, Cance WW, Noel B, Mozingo DW. Game-based mass casualty burn training. Stud Health Technol Inform 2009;142:142-144. [Medline: 19377134]
48. Dev P, Heinrichs WL, Youngblood P. CliniSpace: a multiperson 3D online immersive training environment accessible through a browser. Stud Health Technol Inform 2011;163:173-179. [Medline: 21335784]
49. Jerin JM, Ansell BA, Larsen MP, Cummins RO. Automated external defibrillators: skill maintenance using computer-assisted learning. Acad Emerg Med 1998 Jul;5(7):709-717 [FREE Full text] [doi: 10.1111/j.1553-2712.1998.tb02490.x] [Medline: 9678396]
50. Ponder M, Herbelin B, Molet T, Scherteneib S, Ulicny B, Papagiannakis G, et al. Interactive scenario immersion: Health emergency decision training in JUST project. 2002. URL: http://george.papagiannakis.org/multimedia/130.pdf [accessed 2020-03-02]
51. Laerdal P. MicroSim Prehospital self-directed learning system. Laerdal. 2012. URL: https://www.laerdal.com/gb/docid/12984879/MicroSim-Prehospital [accessed 2020-05-20]
52. Dassault Systemes. StayingAlive3D. iLumens website. 2011. URL: https://3dexperiencelab.3ds.com/en/projects/life/ilumens/ [accessed 2020-05-20]
53. Wattanasoontorn V, Boada I, Sbert M. LISSA: A Serious Game to learn Cardiopulmonary Resuscitation. 2013. URL: https://pdfs.semanticscholar.org/bd7f/697ef12938c4a2e50028edbf54175d0a1eca.pdf [accessed 2020-04-03]
54. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. Comput Methods Programs Biomed 2015 Nov;122(2):282-291. [doi: 10.1016/j.cmpb.2015.08.006] [Medline: 26319184]
55. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Thió-Henestrosa S, Sbert M. 30 : 2: A Game Designed to Promote the Cardiopulmonary Resuscitation Protocol. International Journal of Computer Games Technology 2016;2016:1-14. [doi: 10.1155/2016/8251461]

56. Semeraro F, Frisoli A, Ristagno G, Loconsole C, Marchetti L, Scapigliati A, et al. Relive: a serious game to learn how to save lives. Resuscitation 2014 Jul;85(7):e109-e110. [doi: 10.1016/j.resuscitation.2014.03.306] [Medline: 24709614]
57. Studio Evil. Viva!Game. Viva! Game App website. 2013. URL: https://appadvice.com/app/viva-game/651627940 [accessed 2020-05-25]
58. Schmitz B, Klemke R, Walhout J, Specht M. Attuning a mobile simulation game for school children using a design-based research approach. Computers & Education 2015 Feb;81:35-48. [doi: 10.1016/j.compedu.2014.09.001]
59. Benkhedda S, Bendella F. FASim: A 3D Serious Game for the First Aid Emergency. Simulation & Gaming 2019 Aug 16;50(6):690-710. [doi: 10.1177/1046878119865913]
60. Carvalho J, Horta N, Tome J. Fuzzy Boolean Nets Based Paediatrics First Aid Diagnosis. 2007 Presented at: Annual Meeting of the North American Fuzzy Information Processing Society; 2007; San Diego, CA, USA. [doi: 10.1109/nafips.2007.383916]
61. Saenz Z, Zorrilla A, Zapirain B. Serious Game based on first aid education for individuals with Autism Spectrum Disorder (ASD) using android mobile devices. 2011 Presented at: Conference on Computer Games (CGAMES); 2011; Louisville. [doi: 10.1109/cgames.2011.6000343]
62. Guillén-Nieto V, Aleson-Carbonell M. Serious games and learning effectiveness: The case of It's a Deal!. Computers & Education 2012 Jan;58(1):435-448. [doi: 10.1016/j.compedu.2011.07.015]
63. Zhonggen Y. A Meta-Analysis of Use of Serious Games in Education over a Decade. International Journal of Computer Games Technology 2019 Feb 03;2019:1-8. [doi: 10.1155/2019/4797032]
64. Kebritchi M, Hirumi A. Examining the pedagogical foundations of modern educational computer games. Computers & Education 2008 Dec;51(4):1729-1743. [doi: 10.1016/j.compedu.2008.05.004]
65. O'Neil HF, Wainess R, Baker EL. Classification of learning outcomes: evidence from the computer games literature. Curric. Journal 2005 Dec;16(4):455-474. [doi: 10.1080/09585170500384529]
66. Virvou M, Alepis E. Mobile educational features in authoring tools for personalised tutoring. Computers & Education 2005 Jan;44(1):53-68. [doi: 10.1016/j.compedu.2003.12.020]
67. Torrente J, Borro-Escribano B, Freire M, del Blanco A, Marchiori EJ, Martinez-Ortiz I, et al. Development of Game-Like Simulations for Procedural Knowledge in Healthcare Education. IEEE Trans. Learning Technol 2014 Jan;7(1):69-82. [doi: 10.1109/tlt.2013.35]
68. Dewey J. Experience and Education. In: New York: Simon & Schuster. New York, United States of America: Free Press; 1 edition (July 1 1997); 2018:978-1473.
69. Westera W. Games are motivating, aren't they? Disputing the arguments for digital game-based learning. IJSG 2015 Jun 17;2(2). [doi: 10.17083/ijsg.v2i2.58]
70. Rodríguez A, Boada I, Thió-Henestrosa S, Sbert M. CPRforblind: A video game to introduce cardiopulmonary resuscitation protocol to blind people. Br J Educ Technol 2018 Jul 16;49(4):636-645. [doi: 10.1111/bjet.12627]
71. De Jans S, Van Geit K, Cauberghe V, Hudders L, De Veirman M. Using games to raise awareness: How to co-design serious mini-games? Computers & Education 2017 Jul;110:77-87. [doi: 10.1016/j.compedu.2017.03.009]
72. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati Khodabandeh M, Heidari S, et al. Health Education Serious Games Targeting Health Care Providers, Patients, and Public Health Users: Scoping Review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
73. Glover K, Bodzin A. Learner Analysis to Inform the Design and Development of a Serious Game for Nongaming Female Emerging Health Care Preprofessionals: Qualitative Sample Study. JMIR Serious Games 2020 Feb 06;8(1):e16003 [FREE Full text] [doi: 10.2196/16003] [Medline: 32027312]
74. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
75. Frazer A, Argles D, Wills G. Is Less Actually More? The Usefulness of Educational Mini-games. 2007 Presented at: Seventh IEEE International Conference on Advanced Learning Technologies (ICALT); 2007; Niigata, Japan. [doi: 10.1109/icalt.2007.173]
76. Illanas A, Gallego F, Satorre R, Llorens F. Conceptual mini-games for learning. 2008. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.6252&rep=rep1&type=pdf [accessed 2020-05-02]
77. Jonker V, Wijers M, Galen F. The Motivational Power of Mini-games for the Learning of Mathematics. 2009 Presented at: Proceedings of Third European Conference on Gamebased Learning (ECGBL); 2009; Austria.
78. De Jans S, Van Geit K, Cauberghe V, Hudders L, De Veirman M. Using games to raise awareness: How to co-design serious mini-games? Computers & Education 2017 Jul;110:77-87. [doi: 10.1016/j.compedu.2017.03.009]

Abbreviations

CPR: cardiopulmonary resuscitation

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by M Chover, A Puig; comments to author 12.02.20; revised version received 28.05.20; accepted 11.06.20; published 19.08.20.

Please cite as:
Boada I, Rodriguez Benitez A, Thió-Henestrosa S, Soler J
A Serious Game on the First-Aid Procedure in Choking Scenarios: Design and Evaluation Study
JMIR Serious Games 2020;8(3):e16655
URL: http://games.jmir.org/2020/3/e16655/
doi:10.2196/16655
PMID:32812882

Original Paper

Designing a Virtual Reality Game for Promoting Empathy Toward Patients With Chronic Pain: Feasibility and Usability Study

Xin Tong[1], BSc, MSc; Diane Gromala[1], PhD; Seyedeh Pegah Kiaei Ziabari[1], BSc; Christopher D Shaw[1], PhD

Simon Fraser University, Surrey, BC, Canada

Corresponding Author:
Xin Tong, BSc, MSc
Simon Fraser University
250-13450 102 Ave
Surrey, BC, V3T 0A3
Canada
Phone: 1 7788911232
Email: tongxint@sfu.ca

Abstract

Background: Many researchers have been evaluating how digital media may impact the emotional and perspective taking aspects of empathy in both clinical and nonclinical settings. Despite the growing interest in using virtual reality (VR) and VR games to motivate empathy, few studies have focused on empathy for people who live with chronic pain.

Objective: Chronic pain affects, by conservative estimates, 1 in 5 people in industrialized countries. Despite this prevalence, public awareness of chronic pain was remarkably low until the recent opioid crisis; as a result, stigma remains a problem frequently faced by people who live with this condition. To address this, the VR game *AS IF* was developed to increase nonpatients' empathy toward the growing number of people who live with long-term chronic pain. On the basis of our prior work, we overhauled our approach, designed and built a VR prototype and evaluated it, and offered design suggestions for future research.

Methods: We introduced the design features of the VR game *AS IF* and described the study we devised to evaluate its effectiveness. We adopted a mixed methods approach and compared the empathy-related outcomes in both pre- and posttesting. A total of 19 participants were recruited.

Results: The findings of this study suggest that the VR game was effective in improving implicit and explicit empathy as well as its emotional and perspective taking aspects. More specifically, for the *Empathy Scale*, the total pretest scores (mean 47.33, SD 4.24) and posttest scores (mean 59.22, SD 4.33) did not reach statistical significance (P=.08). However, we did find differences in the subscales. The *kindness* subscale showed a statistically significant increase in the posttest score (mean 15.61, SD 2.85) compared with the pretest score (mean 17.06, SD 2.65;P=.001). For the *Willingness to Help Scale*, a significant increase was observed from a t test analysis (P<.001) of scores before (mean 7.17, SD 2.28) and after (mean 8.33, SD 2.03) the gameplay. The effect size for this analysis was large (d=−1.063).

Conclusions: The contributions of this research are as follows: *AS IF* provides a promising approach for designing VR games to motivate people's empathy toward patients with chronic pain, the study evaluates the potential effectiveness of such a VR approach, and the general design suggestions devised from this study could shed light on future VR game systems.

(JMIR Serious Games 2020;8(3):e17354) doi:10.2196/17354

KEYWORDS

virtual reality; serious games; empathy; chronic pain; game design

Introduction

Background

Pain is a basic and necessary experience that alerts us to physical harm or infection. However, pain is notoriously difficult to describe, and it is harder for one person to understand what another person's pain is like. The International Association for the Study of Pain defined chronic pain as pain that persists for more than 3 months [1], whereas acute pain refers to pain that persists for a shorter time and is expected to subside with healing (eg, pain from surgery or childbirth) [2]. Although most people have experienced acute pain, chronic pain persists well beyond the expected time of healing and can last a lifetime [3]. Although chronic pain is now considered to be a condition [1], its cause and cure have yet to be discovered, and it has no clear biomarkers. In addition, chronic pain often has associated sequelae, from anxiety, depression, insomnia, and cognitive impairment to decreasing mobility [3].

Conservative estimates suggest that 1 in 5 people of all ages live with chronic pain [1,3]. Despite this high prevalence, people who live with chronic pain also face social stigma and high rates of social isolation [4]. More than loneliness, social isolation is correlated with a decreasing quality of life and earlier morbidity [3,5]. As treating this long-term pain means managing it, it is important that caregivers and health professionals find ways to understand the *lived experience* of chronic pain sufferers—in other words, what it is like to actually live with the debilitating effects of long-term pain and how it impacts a patient's biopsychosocial realities, their ability to function, and their quality of life [3].

Definitions of empathy are wide ranging: the most common is "our ability to perceive, understand, and respond to the experiences and behavior of others" [6]. According to Davis [7], empathy has 3 primary dimensions: (1) physical sensations, (2) emotional sensations, and (3) cognitive awareness (perspective taking). In many studies from the medical and cognitive science domains, the emotional and perspective taking dimensions of empathy were the predominant dimensions that were evaluated, along with their effects in both clinical and nonclinical environments [8-13].

However, despite renewed attention on empathy, some studies found that teaching empathy has declined in undergraduate education both in medical schools [8,14,15] and colleges in general [16]. Researchers believe that we can mitigate this situation by teaching empathy in medical schools by *embedding it in students' actual experiences with patients* [17], which can lead to better diagnoses [17-19]. Consequently, empathic communication may help patients manage their health by fostering trust between them and their health professions [9]. Hence, empathy training may similarly improve the attitudes of nonpatients toward patients with chronic pain. (Here, we define *nonpatients* in terms of 2 populations: (1) doctors, nurses, and health professionals who work with patients with chronic pain and (2) caregivers and family members of chronic pain sufferers.)

Previous studies demonstrated that games and video games have a high potential for fostering empathy toward a certain population, even if the players are concerned about winning the game [20-24]. The sense of embodiment in virtual reality (VR) games can appear significantly more effective compared with watching a 2D video [12,25,26]. Evidence also showed that VR games and environments can facilitate a significantly higher level of empathy than videos or traditional media forms [27]. The immersive and convincing nature of VR has profound effects and may confer meaningful benefits for an individual's cognition or behavior [28]. For instance, Bailenson et al [27] found that subjects showed a higher occurrence of thinking from the place of their partners in a VR perspective taking experience. They also observed an improvement in participants' empathy levels toward their partners when they considered the patients' perspective [27]. In other words, two experiential aspects of VR—an immersive sense and an embodied sense—appear to play important roles in empathy [27,29,30].

Objectives

To develop or strengthen a more empathic connection between nonpatients and patients with chronic pain, we designed an interactive VR game entitled *AS IF* [31]. In this VR game, participants *inhabit* an avatar—a 3D character who has chronic pain—from a first-person perspective. Participants were then asked to perform a series of activities and tasks as patients with chronic pain. In other words, *AS IF* was developed to enable participants to not only inhabit the virtual body of a patient with chronic pain but also to metaphorically *walk in the shoes* of this patient by experiencing physical limitations during tasks and to hear that patient's self-talk during each task. We designed *AS IF* with a view to eventually deploying it for clinicians and caregivers. Overall, the goals of this study were to (1) increase empathy toward the growing number of people who live with chronic pain, (2) evaluate the prototype, and (3) offer design suggestions that may guide future research.

Methods

AS IF, the Game Design

Initially, a non-VR desktop version of *AS IF* was developed and tested by the authors using *Microsoft Kinect* [31]. Later, we conducted a study to evaluate the effectiveness of this interactive game system [32]. In that version [31,32], the game tasks involved completing *connect-the-dots* puzzles from a third-person perspective. The player *inhabited* and could move the limbs and head of a virtual avatar—a grandmother living with chronic pain. Players listened to the grandmother's self-talk while solving puzzles; each solved puzzle resulted in the completion of a kitchen task. From analyzing the feedback collected in playtesting, we found that participants were more willing to help people with chronic pain and exhibited positive attitudes about playing from the first-person perspective. However, participants wanted to interact with virtual objects more directly, rather than solving the more abstract connect-the-dots puzzles, and some suggested we design a comparable game in VR for that reason.

Therefore, we redesigned the game and switched from *Microsoft Kinect's* platform to an immersive VR platform, the *HTC VIVE* [33]. In addition, we made significant changes to the game using the *Unity3D* [34] game engine. Here, we outline the revision of *AS IF*.

In this VR version, from the first-person perspective, the player inhabits an avatar of a grandmother who lives with chronic pain. Participants attended to tasks that required physical movements and, as the grandmother, experienced limitations of those movements as she would, a point that was made in her *self-talk* or audible narration. The embodiment aspects of the VR game feature multimodal feedback; design decisions such as this were based on the theory of embodied simulation [35].

Narrative and Stories

Now redesigned for a different medium, a major overhaul involved the visual and interactive aspects of the game.

However, the narrative of *AS IF* primarily remained the same because participants favored the story and settings. Similar to the initial version, the VR game starts with an introductory tutorial that shows players how to use handheld controllers to interact with *AS IF* and its virtual objects. In contrast to the initial version, the movement tasks of *AS IF* are now *realistic*—direct manipulation of objects by using one's hands (via controllers). Instead of connecting the dots to get something done, the VR player makes a cake by direct action, such as breaking an egg. In this new version, participants play from a first-person perspective rather than the prior version's third-person perspective (Figure 1). As the players as grandmother attend to each task of cake making, they simultaneously hear the grandmother's self-talk: the hopes, frustrations, and fears that stem from the ways chronic pain affects her.

Figure 1. The virtual reality game's developer view; and the kitchen in *AS IF*.

Representations of Physical Pain

When the VR player moves each handheld controller, the grandmother avatar (the virtual body inhabited by the player) moves synchronously. A widely adopted *Unity* package named *Final IK* was used to achieve this synchronous movement function [36]. This synchronous movement from a first-person perspective creates an illusion that the player inhabits the avatar—the player thus feels *as if* they are embodied as a grandmother. When the player interacts with the virtual objects, they also experience the avatar's physical limitations that result from pain. The idea is not to induce pain in the player, but rather to enable the player to get a sense *as if* they are a patient with chronic pain. Here, physical *pain* is made *visible* in 2 ways: by limited movement (range and the ability to hold onto an object; Figure 2) and by visual cues. Another visual effect is that when *pain spikes* in the game, red flashes appear to mimic the onset of a headache (Figure 3). Consequently, these indications of pain are a form of feedback, and players quickly learn that they hinder their ability to accomplish tasks through the avatar.

In general, the overhaul of the VR game *AS IF* is distinguished from prior versions designed to facilitate healthy people's empathy toward patients in 2 ways. First, the player can now experience, if not physical pain itself, at least a sense of how pain may limit their ability to move and visual representations (red flashes) of an impending headache. To our knowledge, few VR games use visual representations to depict pain, which may limit physical movement and breakthrough pain. Second, the experience of what it may be like to live with chronic pain is implicitly deployed in the virtual environment—players perform direct, realistic tasks as if she/he is the patient with chronic pain; gain a sense of how pain may limit and interfere with ordinary physical movements; see signals that *something is wrong*; and hear self-talk about the fears and anxiety that living with pain can produce.

Figure 2. Three selected game tasks in *AS IF*: A. mixing the flour with milk; B. adding cream; and C. adding birthday candles.

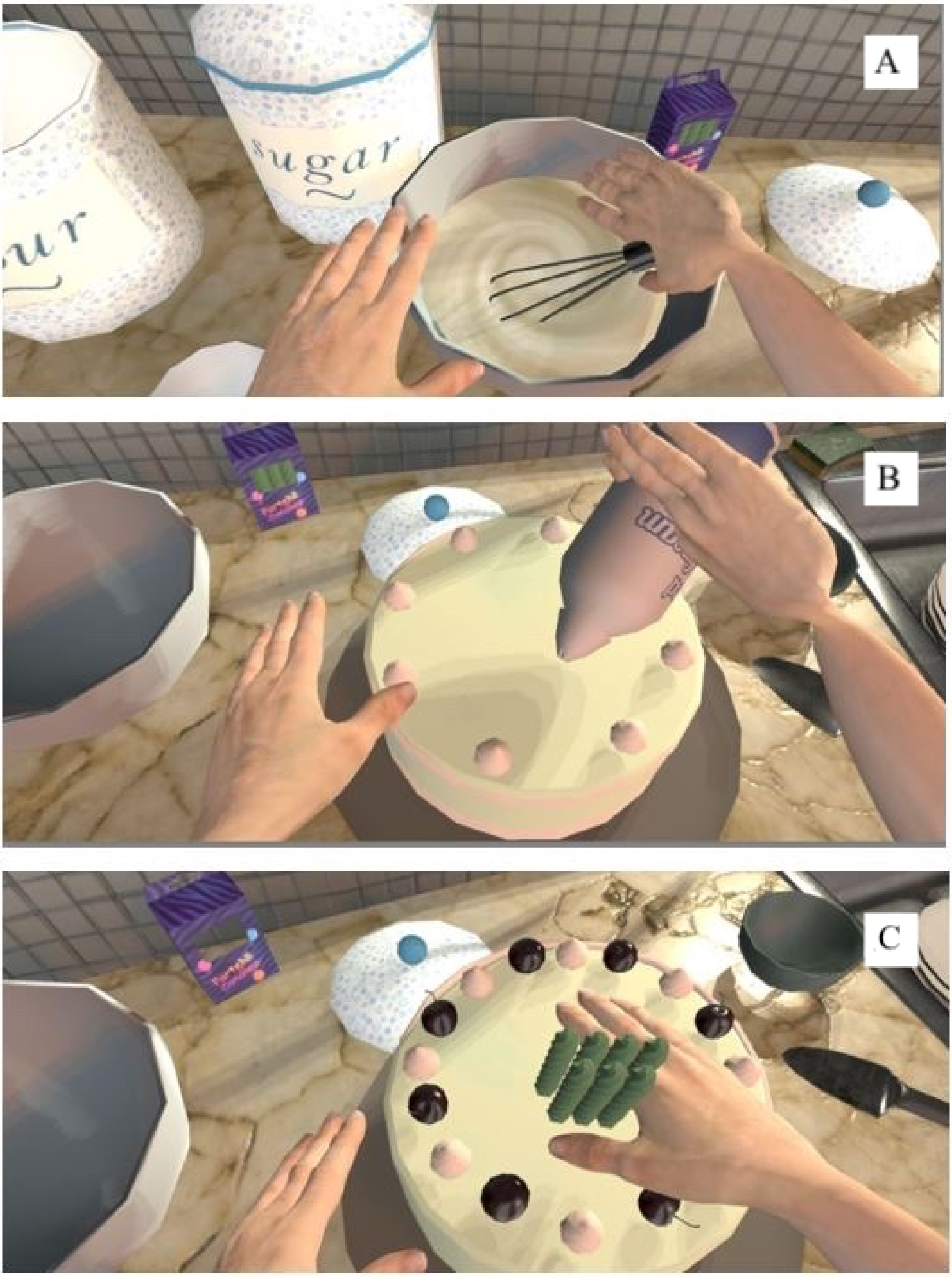


Figure 3. Some of the movements of the virtual avatar's limbs in *AS IF* become limited when pain increases (as circled in the figure). In addition, red flashes signal the onset of a headache. During the flashes, a transparent red layer appears over the visible areas of the virtual environment.

Recruitment

Altogether, 18 participants (4 females), aged 19-39 years (mean 24.8, SD 3.8 years), were recruited through convenience sampling method for this study. We placed advertisements in the university campus media and sent emails to the faculty and student groups. The inclusion criterion was anyone older than 19 years, and the exclusion criteria were anyone who had a pain condition or was a pain patient or did not understand English. No participants had a reported history of a chronic pain diagnosis; this was an essential requirement because personal experience of chronic pain could bias expectations of *AS IF*. Nevertheless, 3 participants reported that they had contact with patients with chronic pain.

Study Intent

The goal of this research study was to determine if a serious VR game such as *AS IF* may influence participants' empathy for patients with chronic pain (raising awareness and fostering positive attitudes toward patients) and to explore what factors may be important to elicit empathy. Furthermore, based on participants' feedback on this version of our design, we summarize the fundamental design principles and game components to help guide future works.

Apparatus

Participants experienced and could interact with objects in the virtual environment via a wired, stereoscopic *HTC VIVE* head-mounted display (HMD) and its handheld controllers. We developed the game using the *Unity3D* game engine, which was responsible for rendering and running the game during the study. These real-time rendered scenes of *AS IF* were sent to the HMD through SteamVR (HTC VIVE) suites; the software was responsible for data between *Unity3D* and the devices via an *API* called *OpenVR*.

Procedures

A mixed-method design approach was adopted in this pre-test, post-test study. Participants' empathy levels towards chronic pain patients before and after playing *AS-IF*, and quantitative questionnaires and qualitative interviews were used in parallel to derive the findings [37]. The study lasted for 35 to 40 min in total. On arrival, participants were briefly introduced to the purpose of the study and the entire procedure and were then asked to read and sign the informed consent form. Next, participants were asked to fill out the preintervention questionnaire, which included the *Empathy Scale* (revised from the *Compassion Scale* [38]), the *Willingness to Help Scale*, the VR-adapted *Other in the Self Scale*, and the *Emotional Wheel* evaluation (Multimedia Appendix 1). During the intervention, participants were first shown how to play *AS IF*, and then, they played it for approximately 10-15 min.

When they finished the game, participants were asked to fill out the posttest questionnaires to assess their level of empathy toward patients with chronic pain. The posttest questionnaires were a repeat of the *Empathy Scale*, the *Willingness to Help Scale*, and the *Emotional Wheel* evaluation. The *Other in the Self Scale* was added to assess the *relationship between self and the first-person perspective* in the VR experience (Multimedia Appendix 1). In addition, participants were given a Sense of Embodiment questionnaire to evaluate their perceived level of immersion in the VR game. Finally, through a 15-min semistructured interview, participants discussed their experience, provided feedback, and offered researchers their ideas about the game. The interview topics were primarily about the game's interactions, depictions of *pain*, and physical impact on participants' empathetic attitudes in *AS IF*. Meanwhile, the session was audio recorded to ensure that the captured data were accurate.

Instruments

We used multiple instruments to measure various aspects of empathy. For instance, the *Empathy Scale* (adapted from

Pommier Compassion Scale [38,39]) measures multiple-dimensioned implicit cognitive changes, whereas the *Willingness to Help Scale* detects explicit cognitive changes. *Wheels of Emotion* reflects emotional changes, and the VR-adapted *Other in the SelfScale* assesses the perspective taking aspect of empathy. As chronic pain is a complex process that involves physical, emotional, and social aspects, we adopted instruments that account for cognitive and emotional perspectives regarding pain (refer to Multimedia Appendix 1 for more details on each scale).

To determine which instruments were validated and the most appropriate, we compared existing instruments, such as the *Basic Empathy Scale* [40], *Toronto Empathy Questionnaires* [41], and Baron-Cohen and Wheelwright's *Empathy Quotient* for people with autism [42]. After thoroughly examining questions from each scale and evaluating how the scales may fit and be adapted for our purposes and population (patients with chronic pain), we selected the *Compassion Scale* [38,39] for its appropriate number and types of questions.

In our previous study of the initial version of *AS IF*, the *Empathy Scale*, the five-point *Numerical Rating Scale* (NRS), and the *Willingness to Help Scale* (10-point NRS) were used to understand if and how that game may have affected participants' cognitive empathy levels. A total of 2 pretest questionnaires were also used in the previous version: (1) the *Empathy Scale* was used to assess implicit empathy (unconscious emotion) and (2) the *Willingness to Help Scale* was used to assess explicit empathy (conscious emotion).

For the new VR version of *AS IF*, the *Wheel of Emotion* [43] and the VR-adapted *Other in the Self Scale* [44] (Figure 4) were adopted to measure any emotional changes associated with empathy and the degree of perspective taking according to the suggestions of Carey et al [45]. The *Wheels of Emotion* provides a comprehensive list of emotions and standardizes the data collected (according to Plutchik's Psycho-Evolutionary Theory of Emotion), whereas the *Other in the Self Scale* helps players articulate how they related to the grandmother's avatar. Moreover, to further investigate embodiment in the VR version, we adapted a *sense of ownership (SoO)* and a *sense of agency (SoA)*, rated on an 11-point NRS (from −5 to 5). The adapted questions are listed in Table 1.

Figure 4. Virtual reality–adapted Other in the Self Scale.

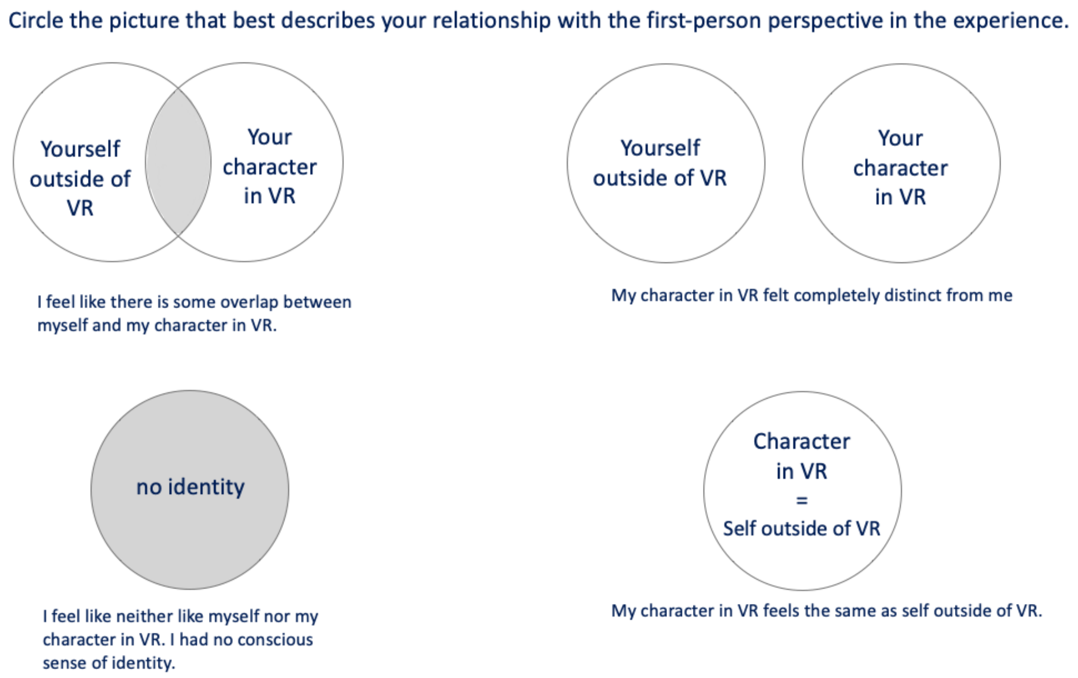


Table 1. The sense of ownership and sense of agency self-reported questionnaires were sorted by categories.

Embodiment aspects	Questions: while playing *AS IF*
Sense of body ownership	• I feel as if the virtual body (eg, arms and hands) were my own (eg, real arms and hands). • I feel my real body (eg, arms and hands) were becoming virtual. • I feel my real body (eg, arms and hands) were moving sometimes.
Sense of body agency	• I feel as if the virtual body (eg, arms and hands) have a will of its own. • I feel the virtual body (eg, arms and hands) would move in the same way as my real body (eg, arms and hands).

Results

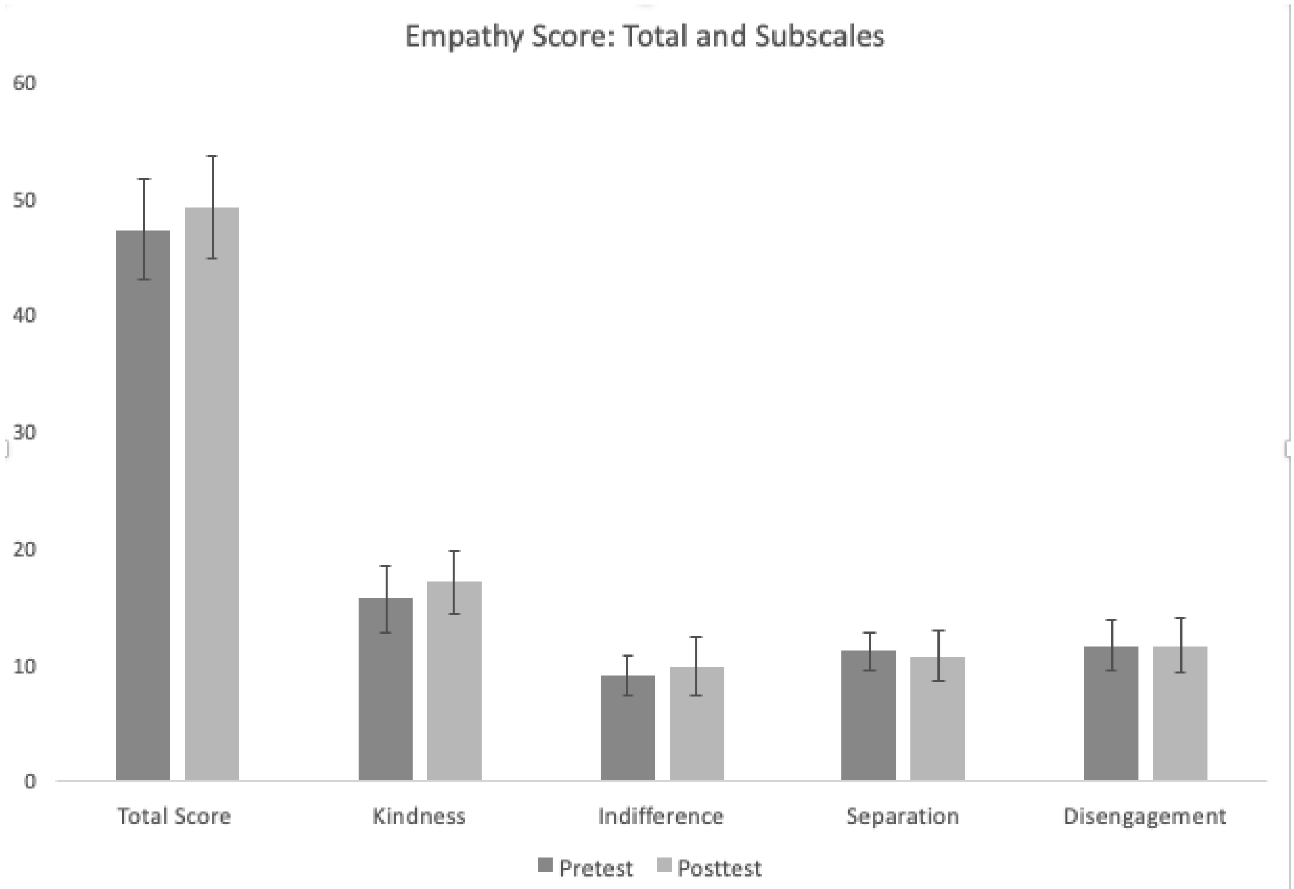
Quantitative Findings

All statistical analyses of the quantitative data were performed using SPSS version 22 (IBM) software [46]. For qualitative data, in-person interviews were first transcribed as electronic textual data. These data were then coded into categories by 2 authors based on preexisting knowledge or hypotheses. After comparing the results, the researchers highlighted significant patterns and summarized and grouped them into themes.

Empathy Scale (Adapted Compassion Scale)

We used a paired *t* test to analyze the differences between the before and after empathy ratings, total scores and subscale scores (Figure 5), kindness, indifference, separation, and disengagement [38]. For the adapted *Empathy Scale*, the total scores from the pretest (mean 47.33, SD 4.24) and the posttest (mean 59.22, SD 4.33) score did not reach statistical significance (t_{17}=−1.41; *P*=.07). However, we found differences in the subscales: the kindness subscale showed a statistically significant increase in the posttest (mean 15.61, SD 2.85) compared with pretest (mean 17.06, SD 2.65; t_{17}=−3.97; *P*=.01). However, indifference (t_{17}=−1.52; *P*=.14), separation (t_{17}=0.75; *P*=.46), and disengagement (t_{17}=0; *P*=.99) subscales were not statistically significant before and after the study. The mean and SD values of separation (pretest: mean 11.11, SD 1.64 and posttest: mean 10.72, SD 2.11), disengagement (pretest: mean 11.61, SD 2.17 and posttest: mean 11.61, SD 2.28), and indifference subscales (pretest: mean 9, SD 1.75 and posttest: mean 9.83, SD 2.46) in pretest and posttest are shown in Figure 5. Admittedly, the scores of the overall empathy and its 3 subscales (indifference, disengagement, and separation), which are different combination of questions from the empathy questionnaire, did not change significantly. However, the statistical significance of the kindness subscale revealed that this aspect of empathy could be potentially altered in a VR game such as *AS IF*.

Figure 5. Mean values from the *Empathy Scale* with error bars before and after the study. From left to right: total score, kindness subscale, indifference subscale, separation subscale and disengagement subscale (y-axis: self-reported ratings from the questions in each category).



Willingness to Help Scale

The *Willingness to Help Scale* has a question that involves a real-world scenario regarding how likely one is to help a patient with chronic pain (refer to Multimedia Appendix 1); it is intended as a means to evaluate the emotional and perspective taking aspects of a participant's empathy for patients with chronic pain. For the scores of the *Willingness to Help Scale* before and after the game intervention and from a *t* test analysis, a significant increase in the posttest score was observed (mean 8.33, SD 2.03) compared with the pretest score (mean 7.17, SD 2.28; t_{17}=−4.51; *P*<.001). The effect size for this analysis was found to be large, according to Cohen convention (*d*=−1.063). This statistically significant increase indicates that the game was able to increase participants' explicit willingness to help people with chronic pain.

The Wheel of Emotions

In their protocol paper for measuring empathy in VR, Carey et al [45] recommended using the *Wheel of EmotionsScale* to measure the emotional aspects of empathy. Specifically, this instrument is intended for understanding emotional empathy or *the spontaneous feeling of oneness with another's emotions* [45]. Therefore, we reported the basic analysis of each participant's emotional changes before and after the study. Overall, 12 of the 18 participants changed from positive emotions (eg, joy, love, and optimism) to negative emotions (eg, sad, helpless, and scared). This may have been influenced by the *AS IF* experience. Six participants reported no changes, regardless of what their initial emotions were. However, half of the 6 participants first described their emotions as negative ones (eg, sad, helpless, and scared). Therefore, most participants' emotions appeared to have changed from a positive to negative direction. Given that the virtual character's self-talk can be characterized by a sense of frustration and fear, these results suggest that *AS IF* fostered emotional empathy.

VR-Adapted Other in the Self Scale

A fourth set of tools was needed to understand cognitive empathy, characterized by understanding another's perspective while also maintaining a distinct sense of self. Regarding the relationship between the virtual avatar and the participant's self, 13 of the 18 participants reported feeling an overlap between their sense of self and the virtual avatar. Three participants felt completely distinct from the avatar, one participant reported not feeling any identity of himself (either inside or outside the game), and one participant reported feeling completely the same. Therefore, 14 of the 19 participants (74%) could relate the virtual avatar to themselves while playing *AS IF.* In general, the results from the *VR-adaptedOther in the Self Scale* suggest that most participants felt the virtual body overlapped with their real identity—the perspective taking aspect of empathy. Therefore, most participants were able to take the perspective of the grandmother who suffers from chronic pain in *AS IF*.

Sense of Embodiment—the SoO and the SoA

As one of the goals was to investigate whether embodiment in VR affected or correlated with changes in empathy, we collected posttest data regarding the participants' SoO (of the avatar) and SoA in VR. On average, participants' scores were higher than zero for both SoO (mean 1.28, SD 2.78) and SoA (mean 1.5, SD 2.65). In the SoO and SoA questionnaires (Table 1) because the rating scale ranged from −5 to 5 (−5 means strongly disagree, and 5 means strongly agree with the statements), the mean values here show that the participants experienced a medium to slightly strong level of body ownership and agency over the virtual avatar.

Correlation Analysis

Pearson correlation tests were also run to test the relationship between SoO and SoA, between the sense of embodiment (comprising SoO and SoA) and the Willingness to Help Scale, and between the sense of embodiment and the *Empathy* Scale (comprising 4 subscales). The results show that SoO is significantly correlated to SoA (r_{18}=0.832; *P*<.001), indicating that the participants' SoO strongly correlates with their SoA in AS IF. Although the correlation between SoO and *Empathy Scale*scores (*P*=.10) and the SoA and Empathy Scale (*P*=.11) did not reach statistical significance, the *P* values fell just short of statistical significance. Finally, the results from the Willingness to Help Scale had statistically significant positive correlations with the kindness subscale (r_{18}=0.632; *P*=.005) and statistically negative correlations with the indifference subscale (r_{18}=−0.531; *P*=.02).

Qualitative Findings

Here, we discuss what the participants thought of *AS IF* and evaluate the strengths and weaknesses of the VR game. We also discuss which of the game's main features may be useful for future research.

Participants' Views About the AS IF VR Game

Although *AS IF* does not simulate the physical feeling of persistent pain, the game achieved its primary goals: to motivate the participants to reflect on experiences of patients with chronic pain and to raise empathy.

One participant said:

> *I think this can help me to understand more about patients. It definitely made me start thinking about how hard other day-to-day tasks would be for people with chronic pain.* [P19]

Overall, participants considered the game interaction to be easy to follow and very intuitive. For instance, P05 said:

> *The interaction was pretty good and very illustrative.*

P08 and P09 said:

> *The interaction is pretty straightforward.*

The 2 approaches of representing pain in VR had pros and cons in providing an immersive experience of a patient with chronic pain. Nonetheless, the VR game shows a high potential for fostering cognitive and emotional empathetic attitudes toward people with chronic pain.

Representing Pain, Approach 1: Restricting Movement to Represent Pain

In general, none of the participants had trouble understanding or completing the VR game's tasks. From the interviews, most participants were aware that the physical limitations imposed in VR represented pain. However, a few participants initially reported that these restrictions felt more like a bug in the program. Overall, approximately one-third of the participants considered the randomly frozen hands/arms to be annoying and more like technical issues. For instance, P08 and P10 told us that:

> *At first, I thought the arms had some delay compared to my real arms, and I thought it was technical difficulties. Then I realized it was the game setting.*

Nonetheless, to a certain degree, this mechanism did achieve the goal because participants' emotional status changed, leading to an increased empathetic attitude toward the grandma patient. Overall, 14 participants reported that their affective changes emerged because of the game and described it using negative emotional words, including *depressed*, *impatient*, *upset*, *frustrated*, *pensive*, *sorry*, and *pity*. For instance, P01, P03, P04, and P11 said:

> *I felt lonely in the game and frustrated while playing the game, but in the end, I am happy to finish the game and achieve the patient's goal [cake making].* [P01]

> *I felt like my movements were slowed down. Plus, I made a mess in the kitchen by dropping things. Emotionally, it was a little discouraging and lowered my confidence with being able to bake all by myself.* [P04]

> *It was pensive. I was thinking like people with chronic pain, how it's gonna be for them.* [P11]

The participants reported that their empathy was elicited when they felt that they were incapable of handling easy daily activities as a virtual patient. P10 said:

> *Suffering pain and I should take a rest and slow down my movement later. I feel that everyone else could make a cake faster than me.*

Although 3 of the participants said they did not have emotional feelings about the grandmother, all agreed that this VR game brought about an awareness. For the first time, participants said they started to think about what life would be like for patients with chronic pain. For instance, P16 said:

> *Although I can't feel any pain, I can feel the difficulty of [the] tasks.*

Interestingly, some participants offered suggestions for improving the game, such as adding more and different forms of feedback regarding movement restrictions, such as a pain meter or digital pain diary. Some participants also suggested that if the granddaughter was visible, she could provide contrast with the physical problems the grandma experiences because of her chronic pain. For instance, P09 said:

> *The limited movement also helps as I cannot move quickly, which is appropriate for a grandma at her age. But, it would be better if there is any contrast, for instance, having a very active child, or people who can move fast.*

Representing Pain, Approach 2: The Red Flashes Signaling the Pain of a Headache

Generally, participants' responses to the red flash effect matched our expectations. Specifically, the red flashes elicited some sense of *pain*, if only vaguely, through a visual effect. In total, participants reported 3 types of sensations when they saw this visual stimulus. The first type reported by the majority of participants was that the reddened world made them feel *pain* and *headaches*.

For example:

> *I did notice the red filter effect, and I can understand that the game was trying to express what the pain patient may have experienced.* [P03]

> *It was very annoying, and I cannot think or move when the red shadow happens.* [P05]

> *I feel very dizzy when the red flashes were coming out, and I cannot think anymore. My brain was entirely blank, and this can definitely represent pain to me.* [P08]

The second type of sensations participants experienced in response to the red flashes were not physical pain per se, but an idea of what the reddened environment was meant to be, and they felt bothered by it. For example, 2 participants mentioned that although they did not feel physiological pain:

> *It can work as an indication to slow down my movement...but I do not feel any physical pain myself.* [P09]

The third type of response to the red flashes was an inability to make an association between the visual indication of an impending headache (or pain spike) with the patient's pain experience. Two participants reported that they did not understand this idea and did not think it represented pain effectively.

The Narrative Strengthened the Sense of Immersion

From participants' feedback, the realistic visual depiction of the kitchen and cake making tasks strengthened the experience of *being* a patient with chronic pain:

> *Everything in this game was so realistic and well-done. I was beginning to embody myself to the character and feel I was there. There were moments that I forgot that was me.* [P07]

In addition to the visual simulation, the audio of the grandmother's inner voice or self-talk also strengthened the sense of immersion and empathy. For instance:

> *It could put me into this situation by narrating that for me.* [P18]

> *Yes, the narrator was expressive, and the voice felt very exhausted and tired. I think the audio was the most influential part and it directed the story.* [P13]

The most significant change in the new VR version of *AS IF* was that tasks are now accomplished directly, rather than indirectly by solving puzzles. In the new VR version, participants complete the tasks of baking a cake by directly interacting with virtual objects, such as stirring together ingredients in a bowl, just as they would in the real world; this also increased the sense of immersion. Moreover, the ordinariness or daily life aspects of the tasks also appear to have succeeded in raising awareness of what life with chronic pain might be like.

As P19 said:

> *Normally, speaking of chronic pain patients, I usually think of the hospital or [them] laying on [a] bed. Baking a cake bring[s] me more awareness about how daily life could be so hard for them too. I won't feel having empathy for them if not doing these tasks. Right now, I feel more related to the grandma in the game.*

Therefore, providing a connection to the virtual avatar—by performing realistic tasks and multimodal sensory input in

VR—better situated the participants *as if* they were *in the patients' shoes*. P11 mentioned that he was thinking of his mother's chronic pain while inhabiting the grandma avatar. He felt frustrated, and his experience in *AS IF* reminded him how hard his mother's life was.

Discussion

Principal Findings

We explored a significant redesign and study results of a VR game, the *AS IF*. It is aimed at motivating *people who do not live with chronic pain* (nonpatients) to better understand the lived experience of chronic pain by increasing empathy. In general, the findings demonstrate that participants had greater degrees of empathy toward patients after playing the VR game. Furthermore, from the semistructured interviews, we were able to gather essential feedback about the strengths and limitations of the current VR design, such as the effectiveness of pain representations. Finally, we extracted critical design issues, implications, and protocol suggestions and offered them to potentially benefit similar research in the future.

Overall, after playing the VR game *AS IF,* participants scored significantly higher on the *Willingness to Help Scale* and the kindness subscale—an adaptation of the *Empathy* questionnaire. These two scores revealed that not only could one VR experience of *AS IF* raise people's awareness of chronic pain but it could also increase their *implicit* and *explicit* empathy. Data from the *kindness* subscale showed implicit cognitive changes, whereas data from the *Willingness to HelpScale* revealed changes in explicit empathic attitudes toward patients with chronic pain. The other three subscales showed nonsignificant differences. Furthermore, the qualitative interview data show that most participants reported that playing this game helped them to understand what a chronic pain patient's life would be like, and that they had never thought about that before. We assume these findings may result from 2 potential reasons: (1) indifference, disengagement, and separation are difficult to affect or change during a single, short period, as in this study, and (2) the design of the VR game *AS IF* focused more on the perspective taking and emotional aspects of empathy, but it did not have specific game features that were meant to increase the four subscales.

The *VR-adaptedOther in the Self Scale* suggested that most participants felt that the virtual body overlapped with their real identity. The findings from the *VR-adapted Other in the Self Scale* also overlapped with the interview results. Some participants said they felt embodied in the grandma avatar who has chronic pain through narrative storytelling, the immersive environment, and the game tasks. Thus, in *AS IF* game, participants were able to understand the perspective of the grandma who has chronic pain.

The sense of embodiment scores showed that, on average, participants could sense owning and controlling the virtual avatar. However, the *Pearson* correlation test revealed no statistical significance between the sense of embodiment in VR (comprising SoO and SoA) and the *Empathy Scale* (posttest) or the sense of embodiment in VR and the *Willingness to Help* scale. We conjecture that 2 reasons might account for this nonsignificant outcome. First, there could be multiple factors that affected empathy levels besides embodiment, such as the narrative and the game's specific tasks (and the *fun/frustration* behind that). Hence, a single factor might not be strong enough to alter overall feelings of empathy. As mentioned in the interviews, participants suggested that tactile feedback might be a better way to indicate pain or the association of visual effects with pain. The second possible explanation could be that the game did not provide a strong enough sense of embodiment to reach statistical significance. In interviews, participants said they wanted the virtual avatar to more closely match their own gender and ethnicity and perhaps even body height and shape. A few participants—a male and a participant whose skin color differs from the avatar's—reported they felt disembodied with the virtual avatar because of its divergent characteristics.

Admittedly, the overall empathy scores and the 3 subscales (indifference, disengagement, and separation) did not change significantly. A crucial issue is how long it takes to affect and change empathy and what factors are important in facilitating such change. For example, implicit empathy may be difficult to change in a short time, in part because of its mental cost [47]. However, in this study, the *VR-adapted Other in the Self* Scale findings suggest that participants identified with the VR avatar, insofar as the avatar was felt to overlap with their real self. Therefore, a perspective taking ability may be critical to being able to influence one's empathetic attitudes toward patients with chronic pain and *painful* experiences. From the qualitative interviews, 2 approaches to representing pain in the virtual body may also facilitate empathy. Although a few participants found the movement restrictions confusing, most reported it made them realize how pain would impact one's range of motion and emotion. Most liked the idea of using the *red flashes* to represent pain spikes and reported that the visual effect felt like a headache or pain (a synesthesia effect of transferring a visual sense to an emotional sense).

Besides facilitating nonpatients' (game players who are health care givers or family members of patients with chronic pain) empathy toward patients with chronic pain, we believe that a VR game such as *AS IF* has the potential to benefit patients and other researchers. Chronic pain experiences are notoriously difficult to describe and are often out of the experiential scope for most people. Hence, VR is one method that may provide clinicians and family members or friends who play a caregiving role with a deeper understanding of a patient's chronic pain condition. VR appears to have the potential to change the player's mind significantly [48] and to stimulate perspective taking and/or behavioral changes that are associated with empathy toward others like patients [49,50]. For instance, in previous research, Platt et al [51] showed how empathic communication, such as clinicians' awareness of the patients' affective states and showing appreciation of the patient's feelings, may reduce patients' feelings of isolation [43]. In addition, the findings, experiences, and design suggestions from this paper may directly benefit other researchers in the future developing empathic games for patients with chronic pain specifically or for patients who must manage similarly invisible chronic conditions. Given our aging population, this may be particularly useful as an approach to medicine shifting from treating acute conditions to managing chronic conditions and promoting wellness.

Some of the participants questioned, “Was my pain sensation the chronic patients’ pain too?” Although it is impossible to make this determination in this study, we assume that the sensations were not the same. For one thing, chronic pain is unique and difficult to describe, let alone recreate specific perception. For another, the interview findings suggested that current gameplay increased the participants’ awareness of chronic pain patients’ situation through the narrative, visual, and audio feedback in *AS IF*.

Limitations

This study also had a few limitations that may affect the empathy outcomes and bring a risk of bias regarding the conclusions. First, the sample size for this preliminary study was small, and a larger number of clinicians will be tested once the prototype is revised and prepared for clinical deployment. Next, to avoid overwhelming participants with assessment instruments, the level of immersion in VR was not measured, so we do not know if there are potential relationships between immersion and changes in empathy. Moreover, we only conducted a preliminary, in-laboratory study using questionnaires and interviews to evaluate changes in empathy, and no real-life assessment has been implemented. However, evaluating the pragmatic aspects of the VR game is definitely something we are planning. This involves running a practical test immediately after the subsequent study, asking participants to donate a portion of their study compensation to a hospital’s foundation or a nonprofit organization for patients with chronic pain. Finally, we did not have a control group and did not conduct a follow-up study to see if any long-term empathy behavioral changes persisted. Investigating these factors to determine whether they affect changes in empathy are important next steps.

Comparison With Prior Work

For a long time, researchers have been looking for evidence about how the sense of embodiment in VR may impact one’s cognitive perception [48]. Various potential impact factors of virtual embodiment on empathy (perspective taking) have been investigated, notably, attitudes toward racial bias [52], gender bias [53], and age [54,55]. We explored the relationship between SoO, SoA, and empathy through correlation tests. However, although participants reported a medium-level SoO and SoA over the virtual body, no significant relationship was discovered.

In summary, to put nonpatients *in the shoes of* patients with chronic pain, players *inhabit* a virtual body of a patient with chronic pain who attends to everyday tasks in the VR game *AS IF*. It simulates several experiences common to chronic pain: physical limitations of movement and a patient’s verbally articulated self-talk. The visual-motor synchronicity of a player’s full-body movements mirrored by the avatar appears to elicit identification with the avatar. The results from the mixed methods study revealed that the game was effective in improving implicit and explicit empathy. Furthermore, the findings showed that the game raised the emotional and perspective taking aspects of players’ empathy. However, no associations were found between the sense of embodiment (SoO and SoA) and the empathy scales in this game. On the basis of the analysis of participants’ feedback, we developed and proposed design suggestions for empathic games designed to facilitate understanding of pain patients. In future work, we plan to iterate the design features and study protocols according to participants’ feedback. We also plan to conduct a randomized controlled study with a larger sample size that is more diverse in terms of gender and age. As for the game, we plan to implement tactile feedback in the controllers (or body sensors) that matches the game tasks, and we would also prepare virtual avatars of different genders and ethnicities.

Acknowledgments

The authors would like to thank all the participants who participated in this study, and the authors gratefully acknowledge the grant from the Natural Sciences and Engineering Research Council of Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaires, interview guide, and correlational tables.
[DOCX File , 775 KB - games_v8i3e17354_app1.docx]

References

1. Chronic Pain Has Arrived in the ICD-11. International Association for the Study of Pain (IASP). URL: https://www.iasp-pain.org/PublicationsNews/NewsDetail.aspx?ItemNumber=8340&navItemNumber=643 [accessed 2019-11-20]
2. Acute Pain. International Association for the Study of Pain (IASP). URL: https://www.iasp-pain.org/GlobalYear/AcutePain [accessed 2019-11-20]
3. Turk D, Monarch E. Psychological Approaches to Pain Management, Second Edition: A Practitioner's Handbook. New York, USA: The Guilford Press; 2018.
4. de Ruddere L, Craig KD. Understanding stigma and chronic pain: a-state-of-the-art review. Pain 2016 Aug;157(8):1607-1610. [doi: 10.1097/j.pain.0000000000000512] [Medline: 26859821]

5. David J, Benjamin A, Baecker R, Gromala D, Birnholtz J. Living with Pain, Staying in Touch: Exploring the Communication Needs of Older Adults with Chronic Pain. In: Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada. [doi: 10.1145/1979742.1979751]
6. Szanto T, Krueger J. Introduction: empathy, shared emotions, and social identity. Topoi 2019 Feb 26;38(1):153-162. [doi: 10.1007/s11245-019-09641-w]
7. Davis M. Empathy: A Social Psychological Approach. Boulder, CO, US: Westview Press; 1994.
8. Quince T, Thiemann P, Benson J, Hyde S. Undergraduate medical students' empathy: current perspectives. Adv Med Educ Pract 2016;7:443-455 [FREE Full text] [doi: 10.2147/AMEP.S76800] [Medline: 27536168]
9. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. J Med Libr Assoc 2018 Oct;106(4):498-500 [FREE Full text] [doi: 10.5195/jmla.2018.518] [Medline: 30271295]
10. Bailenson J. Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do. New York, USA: W W Norton & Company; 2018.
11. van Loon A, Bailenson J, Zaki J, Bostick J, Willer R. Virtual reality perspective-taking increases cognitive empathy for specific others. PLoS One 2018;13(8):e0202442 [FREE Full text] [doi: 10.1371/journal.pone.0202442] [Medline: 30161144]
12. Shin D. Empathy and embodied experience in virtual environment: to what extent can virtual reality stimulate empathy and embodied experience? Comput Hum Behav 2018 Jan;78:64-73. [doi: 10.1016/j.chb.2017.09.012]
13. Banakou D, Slater M. Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. Sci Rep 2017 Oct 27;7(1):14227 [FREE Full text] [doi: 10.1038/s41598-017-14620-5] [Medline: 29079802]
14. Mirani S, Shaikh N, Tahir A. Assessment of clinical empathy among medical students using the Jefferson scale of empathy-student version. Cureus 2019 Feb 28;11(2):e4160 [FREE Full text] [doi: 10.7759/cureus.4160] [Medline: 31058043]
15. Pagano A, Robinson K, Ricketts C, Jones J, Henderson L, Cartwright W, et al. Empathy Levels in Canadian Paramedic Students: A Longitudinal Study. FIRST: Fanshawe Innovation Research Scholarship Teaching. 2019. URL: https://first.fanshawec.ca/fhcsps_publicsafety_facultystaffpublications/23 [accessed 2020-01-22]
16. Empathy: College Students Don't Have as Much as They Used to. University of Michigan News. 2010. URL: https://news.umich.edu/empathy-college-students-don-t-have-as-much-as-they-used-to/ [accessed 2019-09-08]
17. Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med 2011 Aug;86(8):996-1009. [doi: 10.1097/ACM.0b013e318221e615] [Medline: 21670661]
18. Reynolds W. The Measurement and Development of Empathy in Nursing. New York, USA: Routledge; 2017.
19. Percy M, Richardson C. Introducing nursing practice to student nurses: how can we promote care compassion and empathy. Nurse Educ Pract 2018 Mar;29:200-205. [doi: 10.1016/j.nepr.2018.01.008] [Medline: 29427943]
20. Konrath SH, O'Brien EH, Hsing C. Changes in dispositional empathy in American college students over time: a meta-analysis. Pers Soc Psychol Rev 2011 May;15(2):180-198. [doi: 10.1177/1088868310377395] [Medline: 20688954]
21. Kral TR, Stodola DE, Birn RM, Mumford JA, Solis E, Flook L, et al. Neural correlates of video game empathy training in adolescents: a randomized trial. NPJ Sci Learn 2018;3:13 [FREE Full text] [doi: 10.1038/s41539-018-0029-6] [Medline: 30631474]
22. Lemmings (Video Game). Wikipedia. 2019. URL: https://en.wikipedia.org/w/index.php?title=Lemmings_(video_game)&oldid=914380596 [accessed 2019-09-20]
23. Greitemeyer T, Osswald S, Brauer M. Playing prosocial video games increases empathy and decreases schadenfreude. Emotion 2010 Dec;10(6):796-802. [doi: 10.1037/a0020194] [Medline: 21171755]
24. Olivier L, Sterkenburg P, van Rensburg E. The effect of a serious game on empathy and prejudice of psychology students towards persons with disabilities. Afr J Disabil 2019;8:328 [FREE Full text] [doi: 10.4102/ajod.v8i0.328] [Medline: 31049304]
25. Freina L, Canessa A. Immersive vs Desktop Virtual Reality in Game Based Learning. CiteSeerX. URL: http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=7A68CDCE1F13C71E16070BF8FABA19EF?doi=10.1.1.727.9647 [accessed 2019-09-09]
26. Perez-Marcos D. Virtual reality experiences, embodiment, videogames and their dimensions in neurorehabilitation. J Neuroeng Rehabil 2018 Nov 26;15(1):113 [FREE Full text] [doi: 10.1186/s12984-018-0461-0] [Medline: 30477527]
27. Herrera F, Bailenson J, Weisz E, Ogle E, Zaki J. Building long-term empathy: a large-scale comparison of traditional and virtual reality perspective-taking. PLoS One 2018;13(10):e0204494 [FREE Full text] [doi: 10.1371/journal.pone.0204494] [Medline: 30332407]
28. Slater M, Sanchez-Vives MV. Enhancing our lives with immersive virtual reality. Front Robot AI 2016 Dec 19;3:19. [doi: 10.3389/frobt.2016.00074]
29. Lim C, Berry A, Hirsch T, Hartzler A, Wagner E, Ludman E, et al. 'It just seems outside my health': how patients with chronic conditions perceive communication boundaries with providers. DIS (Des Interact Syst Conf) 2016 Jun;2016:1172-1184 [FREE Full text] [doi: 10.1145/2901790.2901866] [Medline: 28804790]
30. Boltz LO, Henriksen D, Mishra P. Rethinking technology & creativity in the 21st century: empathy through gaming - perspective taking in a complex world. Techtrends Tech Trends 2015 Oct 15;59(6):3-8. [doi: 10.1007/s11528-015-0895-1]

31. Jin W, Ulas S, Tong X. AS IF: A Game as an Empathy Tool for Experiencing the Activity Limitations of Chronic Pain Patients Share on. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2851581.2890369]
32. Tong X, Ulas S, Jin W, Gromala D, Shaw C. The Design and Evaluation of a Body-Sensing Video Game to Foster Empathy Towards Chronic Pain Patients. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth'17; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154869]
33. Vive. URL: https://www.vive.com/us/ [accessed 2019-01-15]
34. Unity Real-Time Development Platform. URL: https://unity.com/frontpage [accessed 2019-05-21]
35. Gallese V, Sinigaglia C. What is so special about embodied simulation? Trends Cogn Sci 2011 Nov;15(11):512-519. [doi: 10.1016/j.tics.2011.09.003] [Medline: 21983148]
36. The Final Inverse Kinematics Solution for Unity. Unity Asset Store - Best Assets for Game Making . URL: https://assetstore.unity.com/packages/tools/animation/final-ik-14290 [accessed 2019-05-23]
37. Creswell J, Clark V. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage Publications; 2011.
38. Elizabeth AP. Copyright. Self-Compassion. 2011. URL: http://self-compassion.org/wp-content/uploads/2015/02/POMMIER-DISSERTATION-compassion.pdf [accessed 2016-09-02]
39. Pommier E, Neff KD, Tóth-Király I. The development and validation of the compassion scale. Assessment 2020 Jan;27(1):21-39. [doi: 10.1177/1073191119874108] [Medline: 31516024]
40. Basic Empathy Scale (BES). The RAND Corporation. URL: https://www.rand.org/education-and-labor/projects/assessments/tool/2006/basic-empathy-scale-bes.html [accessed 2020-01-22]
41. Spreng RN, McKinnon MC, Mar RA, Levine B. The Toronto empathy questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. J Pers Assess 2009 Jan;91(1):62-71 [FREE Full text] [doi: 10.1080/00223890802484381]
42. Baron-Cohen S, Wheelwright S. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. J Autism Dev Disord 2004 Apr;34(2):163-175. [doi: 10.1023/b:jadd.0000022607.19833.00] [Medline: 15162935]
43. Halim A, Foster A, Ayala L, Musser E. The physiological nature of caring: understanding nonverbal behavior. In: Foster AE, Yaseen ZS, editors. Teaching Empathy in Healthcare: Building a New Core Competency. New York, USA: Springer International Publishing; 2019.
44. Weidler D, Clark E. A distinct association: inclusion of other in the self and self-disclosure. New School Psychology Bulletin 2011;9(1):- [FREE Full text] [doi: 10.1037/e505232012-005]
45. Carey K, Saltz E, Rosenbloom J, Micheli M, Choi J, Hammer J. Toward Measuring Empathy in Virtual Reality. In: Extended Abstracts: Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY'17; October 15-18, 2017; Amsterdam, Netherlands. [doi: 10.1145/3130859.3131325]
46. SPSS Software. IBM. URL: https://www.ibm.com/analytics/spss-statistics-software [accessed 2020-01-23]
47. Cameron CD, Hutcherson CA, Ferguson AM, Scheffer JA, Hadjiandreou E, Inzlicht M. Empathy is hard work: people choose to avoid empathy because of its cognitive costs. J Exp Psychol Gen 2019 Jun;148(6):962-976. [doi: 10.1037/xge0000595] [Medline: 30998038]
48. Bailey JO, Bailenson JN, Casasanto D. When does virtual embodiment change our minds? Presence-Teleop Virt 2016 Dec;25(3):222-233. [doi: 10.1162/pres_a_00263]
49. Longo MR, Schüür F, Kammers MP, Tsakiris M, Haggard P. What is embodiment? A psychometric approach. Cognition 2008 Jun;107(3):978-998. [doi: 10.1016/j.cognition.2007.12.004] [Medline: 18262508]
50. Schäfer G, Prkachin KM, Kaseweter KA, Williams AC. Health care providers' judgments in chronic pain: the influence of gender and trustworthiness. Pain 2016 Aug;157(8):1618-1625. [doi: 10.1097/j.pain.0000000000000536] [Medline: 26934512]
51. Platt FW, Platt CM. Empathy: a miracle or nothing at all? J Clin Outcomes Manag 2011 Jan;18(1):27-33 [FREE Full text]
52. Behm-Morawitz E, Pennell H, Speno AG. The effects of virtual racial embodiment in a gaming app on reducing prejudice. Commun Monogr 2016 Jan 14;83(3):396-418. [doi: 10.1080/03637751.2015.1128556]
53. Lopez S, Yang Y, Beltran K, Kim S, Cruz HJ, Simran C, et al. Investigating Implicit Gender Bias and Embodiment of White Males in Virtual Reality with Full Body Visuomotor Synchrony. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300787]
54. Banakou D, Groten R, Slater M. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. Proc Natl Acad Sci U S A 2013 Jul 30;110(31):12846-12851 [FREE Full text] [doi: 10.1073/pnas.1306779110] [Medline: 23858436]
55. Tajadura-Jiménez A, Banakou D, Bianchi-Berthouze N, Slater M. Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. Sci Rep 2017 Aug 29;7(1):9637 [FREE Full text] [doi: 10.1038/s41598-017-09497-3] [Medline: 28851953]

Abbreviations

HMD: head-mounted display
NRS: Numerical Rating Scale
SoA: sense of agency
SoO: sense of ownership
VR: virtual reality

Edited by G Eysenbach; submitted 09.12.19; peer-reviewed by L Cilar, Y Lin; comments to author 20.01.20; revised version received 01.04.20; accepted 03.06.20; published 07.08.20.

Please cite as:
Tong X, Gromala D, Kiaei Ziabari SP, Shaw CD
Designing a Virtual Reality Game for Promoting Empathy Toward Patients With Chronic Pain: Feasibility and Usability Study
JMIR Serious Games 2020;8(3):e17354
URL: http://games.jmir.org/2020/3/e17354/
doi:10.2196/17354
PMID:32763883

Original Paper

Results and Guidelines From a Repeated-Measures Design Experiment Comparing Standing and Seated Full-Body Gesture-Based Immersive Virtual Reality Exergames: Within-Subjects Evaluation

Wenge Xu[1], BSc, MSc; Hai-Ning Liang[1], BA, PhD; Qiuyu He[1], BSc; Xiang Li[1]; Kangyou Yu[1]; Yuzheng Chen[1]

Xi'an Jiaotong-Liverpool University, Suzhou, China

Corresponding Author:
Hai-Ning Liang, BA, PhD
Xi'an Jiaotong-Liverpool University
111 Ren'ai Road
Suzhou
China
Phone: 86 51288161516
Email: haining.liang@xjtlu.edu.cn

Abstract

Background: Although full-body seated exercises have been studied in a wide range of settings (ie, homes, hospitals, and daycare centers), they have rarely been converted to seated exergames. In addition, there is an increasing number of studies on immersive virtual reality (iVR) full-body gesture-based standing exergames, but the suitability and usefulness of seated exergames remain largely unexplored.

Objective: This study aimed to evaluate the difference between playing a full-body gesture-based iVR standing exergame and seated exergame in terms of gameplay performance, intrinsic motivation, and motion sickness.

Methods: A total of 52 participants completed the experiment. The order of the game mode (standing and sitting) was counterbalanced. Gameplay performance was evaluated by action or gesture completion time and the number of missed gestures. Exertion was measured by the average heart rate (HR) percentage (AvgHR%), increased HR%, calories burned, and the Borg 6-20 questionnaire. Intrinsic motivation was assessed with the Intrinsic Motivation Inventory (IMI), whereas motion sickness was assessed via the Motion Sickness Assessment Questionnaire (MSAQ). In addition, we measured the fear of falling using a 10-point Likert scale questionnaire.

Results: Players missed more gestures in the seated exergame than in the standing exergame, but the overall miss rate was low (2.3/120, 1.9%). The analysis yielded significantly higher AvgHR%, increased HR%, calories burned, and Borg 6-20 rating of perceived exertion values for the seated exergame (all P<.001). The seated exergame was rated significantly higher on peripheral sickness (P=.02) and sopite-related sickness (MSAQ) (P=.004) than the standing exergame. The score of the subscale "value/usefulness" from IMI was reported to be higher for the seated exergame than the standing exergame. There was no significant difference between the seated exergame and standing exergame in terms of intrinsic motivation (interest/enjoyment, P=.96; perceived competence, P=.26; pressure/tension, P=.42) and the fear of falling (P=.25).

Conclusions: Seated iVR full-body gesture-based exergames can be valuable complements to standing exergames. Seated exergames have the potential to lead to higher exertion, provide higher value to players, and be more applicable in small spaces compared with standing exergames. However, gestures for seated exergames need to be designed carefully to minimize motion sickness, and more time should be given to users to perform gestures in seated exergames compared with standing exergames.

(JMIR Serious Games 2020;8(3):e17972) doi:10.2196/17972

KEYWORDS

exergames; immersive virtual reality; standing exergame; seated exergame; exercising

Introduction

Background

Physical inactivity has been identified as the fourth leading cause of death globally [1]. It is now well established that a sedentary lifestyle is a unique risk factor for several illnesses, such as type 2 diabetes and cardiovascular diseases, which account for about 30% of global mortality [2]. Exergames represent a promising approach that has been widely examined for various population groups (ie, children [3], young individuals [4], and older adults [5]) to promote regular exercise (defined as planned, structured, repetitive, and intentional movement intended to improve or maintain physical fitness) in unmotivated or inactive target groups [6,7].

In recent years, exergames have been proven to have the potential to improve enjoyment, motivation, and long-term engagement when compared with other conventional exercises (eg, cardiovascular exercises like biking [8,9]), and as such, they can be effective in promoting both physical and mental health [10,11]. Various nonimmersive virtual reality (VR) [12] (like using interfaces such as a flat-screen television/monitor) exergames have been designed to encourage people to be more active [5] and promote a positive lifestyle [13] and self-care [14]. Previous literature has shown that exergames could bring physical and health outcomes to players. For example, Peng et al [15] performed a meta-analysis of energy expenditure in exergames, and their main finding suggests that exergames are as effective as traditional exercises in facilitating light- and moderate-intensity physical exertion. Huang et al [16] reported that participants who were enthusiastic about exercising showed positive changes in happiness, perceived energy levels, and relaxation in a 2-week exergame intervention. Sapi et al [17] reported that participants showed improvements in balance following a 6-week exergame intervention, and the improvements were in favor of using the exergame than conventional balance training. da Silva Alves et al [18] found that participants showed improvements in functional well-being and physical well-being after 10 sessions of exergaming. In the study by Garcia et al [19], participants showed improvements in stepping, standing balance, gait speed, and mobility following a 12-week exergame intervention.

With the recent advances in immersive virtual reality (iVR) head-mounted displays (HMDs), an increasing number of iVR exergames [20-22] are being developed. They have opened the possibility of altering more radically how we engage users in performing exercises. Studies have shown that iVR has the potential to produce benefits that other types of displays (ie, a standard display like television) cannot offer. For instance, iVR exergames can offer users the illusion of more exceptional physical capabilities than they have. As such, iVR may increase motivation for exercising in general [23]. Moreover, iVR games can offer benefits such as increased perceived competence and the feeling of body movements that are more in line with how we perform exercise in the real world. Participants have described exaggerated movements to be natural, fun, and empowering [24]. Furthermore, exercising in iVR has been found to be an effective intervention to increase enjoyment and motivation than standard televisions or monitors [8,25], where enjoyment and motivation are, in turn, linked to increased adherence to physical exercises in general [26-28].

Full-body gesture-based exergames have been widely explored with people in standing positions in iVR [20,22,23]. However, they have not been adapted and explored in seated versions. Seated iVR exergames could have the following benefits: (1) suitability for users with a sedentary lifestyle (eg, university students [29]); (2) feasibility for mobility-impaired users (eg, elderly users and wheelchair users [5,30]); (3) possibility of reducing the risk of injuries due to falls or motion sickness [31,32]; and (4) avoidance of injury from hitting other objects (eg, furniture) when the space is small or surroundings are cluttered, because players are not required to walk around.

Motivation can be divided into intrinsic (enjoyment of the activity) and extrinsic (driven by external outcomes, eg, losing weight and improving fitness) [33]. Intrinsic motivation (ie, motivation derived from enjoyment and satisfaction gained from an activity) plays an essential role in long-term adherence to exercising [26,27,33], whereas extrinsic motivation, such as competitive pressure, may lead to tension and feelings of compulsion, and can diminish intrinsic motivation [34,35]. There is evidence that exergames increase enjoyment and intrinsic motivation compared with conventional exercises (eg, biking) and distract from uncomfortable bodily sensations [25,36-39].

A sedentary lifestyle is a problem for older adults and people with physical disabilities and is a serious health problem among university students [40]. Most of the research on exergames has been targeted at older adults or disabled people [5,30,41] but not university students, who are underrepresented in such studies. Research has shown that lack of time and not liking exercising are the major barriers for university students [29]. These barriers could be overcome by using full-body gesture-based exergames that can be played either standing or seated at any time and in small spaces because exergames are perceived to be more enjoyable and preferred by university students than other conventional exercises (eg, cardiovascular exercises like biking [8,9]).

Goal of the Study

The focus of this research was to evaluate the playability and user experience of a seated iVR exergame compared with a similar standing exergame among university students in terms of gameplay performance (ie, action completion time and number of gestures missed) and user experience (ie, motion sickness, intrinsic motivation, and fear of falling).

Methods

Experiment Design

We employed a one-way within-subjects experiment design where the independent variable was *game mode* with two levels (standing and sitting). The order of the game mode was counterbalanced to compensate for any learning effects. The whole experiment lasted between 30 and 40 minutes for each participant depending on their tiredness level and resting heart rate (RestHR).

Participants

Participants were recruited from a local university campus through posters, social media platforms, and a mailing list. The study included university students with the ability to speak English, who were not disabled, were not pregnant (because of the physical exertion required to play the game), and had not consumed any alcohol during the day (blood alcohol level of approximately 0.07% could reduce symptoms of cybersickness [42], which might affect the results of our study).

Participants were excluded from the experiment if they (1) answered "yes" to any of the Physical Activity Readiness Questionnaire (PAR-Q) [43] questions, (2) had a resting blood pressure higher than 140/90 mmHg, and (3) had a RestHR level that was too low (ie, RestHR <62 beats/min for a 16 to 19-year-old female, RestHR <60 beats/min for a 20 to 39-year-old female, RestHR <56 beats/min for a 16 to 19-year-old male, or RestHR <55 beats/min for a 20 to 39-year-old male) or too high (ie, RestHR >94 beats/min for a 16 to 19-year-old female, RestHR >89 beats/min for a 20 to 39-year-old female, RestHR >87 beats/min for a 16 to 19-year-old male, or RestHR >84 beats/min for a 20 to 39-year-old male) [44].

All participants received drinks and snacks for their participation after they finished the experiment. The University Ethics Committee at Xi'an Jiaotong-Liverpool University approved the experiment. All participants signed informed consent forms prior to taking part in the study, and the research protocol was approved by the University Ethics Committee.

To determine the sample size for the study, a statistical power analysis was performed. This statistical power analysis was not based on data from prior studies owing to limited similar work. It was based on general considerations about the trade-offs between the ability to detect certain effects and the feasibility to acquire a sufficiently large sample. We used a sample size calculation software program (G*Power version 3.1.9.2 for Windows), with an effect size of 0.5 (Cohen *d*), statistical power of .90, and statistical level of significance of .05. The sample size was established at 44 participants, and we decided to recruit an additional 10 participants in the case of dropouts.

Game

The game was implemented in Unity3D with the SteamVR plugin to enable positional tracking of the HTC VIVE trackers and controllers.

Rules and Logic

A game is an activity that requires at least one player, has rules, and has a victory condition [45]. The design of our game followed this definition and was inspired by the game called *Just Dance*, which requires users to follow and imitate the dancing gestures one by one and has been used in some prior studies [46]. However, the level of our program consists of a sequence of exercise gestures instead of dance movements. As the game starts, the player needs to follow the body gestures of the instructor in the VR system to move his/her body accordingly. For ease of reproducing the gestures, a gesture is deemed completed if specific joint positions (eg, head, controllers, and trackers) of the player meet the predefined variables of corresponding gestures based on a simple rule-based system and if the player can keep the pose for 0.4 seconds, which was determined from the results of a pilot study with 10 participants, where we found that a short pose hold time could lead to gesture misrecognitions and a long pose hold time could lead to player fatigue easily. In addition, this time of 0.4 seconds was informed from the literature in other fields (eg, text entry [47]). A badge [48] is given to players when they complete every 10 actions as an in-game achievement to motivate them to follow and replicate the gestures carefully. The victory condition was to successfully follow the instructor's gestures and not fail to follow these gestures three times in a row. In addition, our game warned users when they were not paying attention to the virtual instructor by tracking the rotation data from the HMD. Both visual and auditory feedback were provided to encourage players to continue playing.

Game Procedure

The game starts with a calibration phase (Figure 1A) for the system to take into account the individual differences of players. The player needs to lift the hands midair and confirm having finished the calibration by pressing a button on the controllers. The system then records the position data of the head, hands, and feet. After the calibration phase, the game progresses to the training (warm-up) phase (Figure 1B), where the player needs to follow the virtual instructor to perform two rounds of six gestures with a fixed order to become familiar with the gestures that need to be performed. The gameplay phase (Figure 1C) starts after the training phase, where the player needs to follow the virtual instructor who performs gestures presented in a random manner.

Figure 1. (A) Calibration, (B) training, and (C) game phases for the seated version; the process for the standing version is the same, except the instructor is standing instead of sitting.

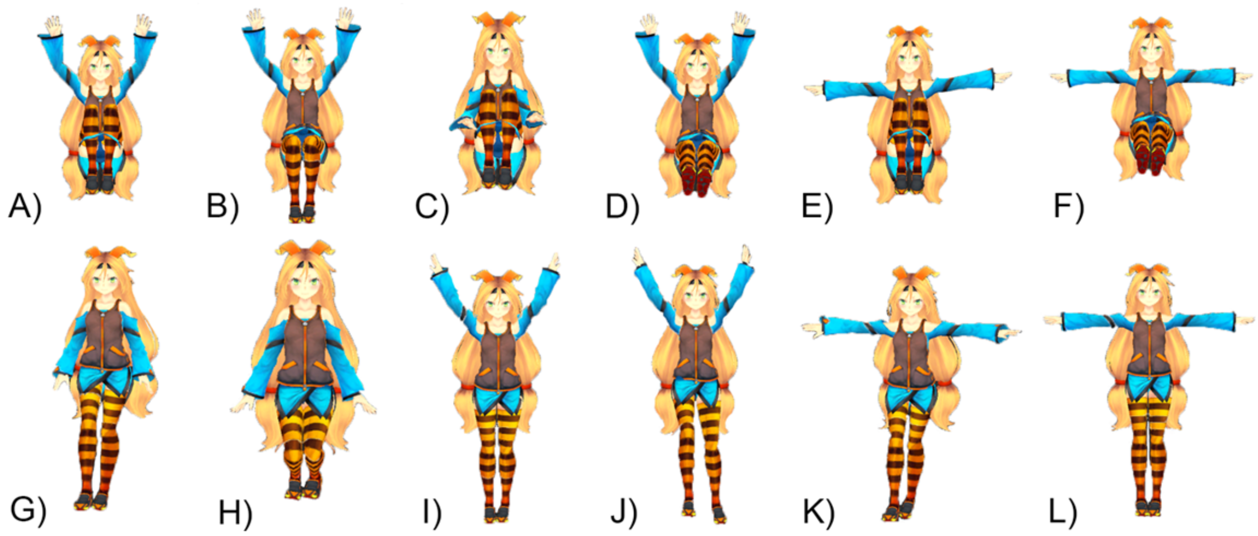
Included Gestures

Table 1 shows the intensity of the gestures selected for our game. Figure 2 shows the final pose of the gestures involved in the sitting and standing versions of the exergame. The selection process of gestures and the intensity values were informed by the results of a pilot study (Multimedia Appendix 1).

Table 1. Intensity level of the gestures used in the experiment.

Gesture	Intensity (%)
Sitting gesture	
Hands up	32.30
Knees up	32.27
Feet up + hands up	35.10
Feet up + hands stretched	34.75
Knees up + hands up	42.39
Knees up + hands stretched	44.03
Standing gesture	
Hands up	31.00
Hands stretched	37.20
Left/right kick	27.25
Squat	50.69
Hand stretched + kick	43.88
Hands up + kick	46.05

Figure 2. Seated gestures: (A) Hands up + knees up; (B) Hands up; (C) Knees up; (D) Feet up + hands up; (E) Hands stretched + knees up; (F) Hands stretched + feet up. Standing gestures: (G) Left/right kick; (H) Squat; (I) Hands up; (J) Left/right kick + hands up; (K) Left/right kick + hands stretched; (L) Hands stretched.



Outcome Measurements

Exertion

We measured participants' exertion based on HR and calories burned using a Polar OH1 wrist-strap monitor. Average HR (AvgHR%) was expressed as a percentage of a participant's estimated maximum HR (MaxHR), where MaxHR was estimated as 220 minus age [49]. This measure is commonly used in exercise studies to confirm that participants are working at a required level of exertion. Additionally, we measured the increased HR%, which was the difference between the HR% at the beginning and the end of the game phase, for a direct comparison of both versions. Calories burned were calculated using the Polar Beat mobile app with the activity set as other indoor activity and the user profile of the app calibrated to each participant. We started recording the HR and calories burned as soon as the participants finished the training phase. Furthermore, the Borg CR 6-20 [50] rating of perceived exertion (RPE) was used to measure the participants' perceived exertion level. It describes the physical efforts involved in completing the game as perceived by the participants, with "6" indicating "no exertion," "13" indicating "somewhat hard exertion," and "20" indicating "maximum exertion." Borg RPE is frequently used in exercise sciences as a quantitative measure of perceived exertion when exercising [50,51], and it has been applied to studies with iVR exergames [4,20].

Gameplay Data

We collected several types of data in the background, including (1) the action completion time of every successfully performed gesture, which was equal to the time spent by the user to perform the same pose and hold the pose for 0.4 seconds, (2) the number of gestures missed for each gesture type, and (3) the real-time HR data from the Polar OH1 optical HR sensor for every 0.2 seconds in the actual experiment stage. Therefore, we analyzed (1) the average action completion time, (2) the total number of missed gestures, and (3) the plot profile of real-time HR.

Motion Sickness

Motion sickness was assessed using the self-reported 16-item Motion Sickness Assessment Questionnaire (MSAQ) [52], which is a valid descriptor of motion sickness in the general population that covers the following four dimensions of motion sickness: gastrointestinal (stomach sick, queasy, nauseated, and vomit), central (faint-like, lightheaded, dizzy, spinning, and disoriented), peripheral (sweaty, clammy, and hot/warm), and sopite-related (annoyed, drowsy, tired, and uneasy). The results from the MSAQ were correlated strongly with the overall scores from the Pensacola diagnostic index (r=0.81; P<.001) and the nausea profile (r=0.92; P<.001) [52]. It has been found that the MSAQ is a valid evaluation tool and that it is advantageous to use this multidimensional questionnaire rather than the one-dimensional form [52]. The questionnaire has been widely used in studies dealing with virtual environments [53-55]. The scale ranges from 1 (not at all) to 9 (severely). A lower score is associated with lower motion sickness.

Intrinsic Motivation

Intrinsic motivation was measured using the self-reported 25-item version of the Intrinsic Motivation Inventory (IMI) [56], which covers the following four subscales: interest/enjoyment, perceived competence, pressure/tension, and value/usefulness. Although IMI includes seven subscales, only interest/enjoyment measures intrinsic motivation and is considered the primary self-reporting measure. We included the perceived competence and pressure/tension subscales because they are positive and negative predictors of intrinsic motivation, respectively. In addition, the subscale value/usefulness has been used in internalization studies [57] and can provide us with an idea about how people internalize and self-regulate themselves with respect to the activities that they experience as useful or valuable for themselves. The IMI has gained widespread acceptance as a multidimensional measure of intrinsic motivation in the context of sports and exercising [58,59] and has been widely used in studies dealing with iVR exergames [23,60,61]. Each item was rated on a severity scale ranging from 1 (not at all) to 7 (very). A higher score indicates a more internally motivated self-regulated physical activity behavior.

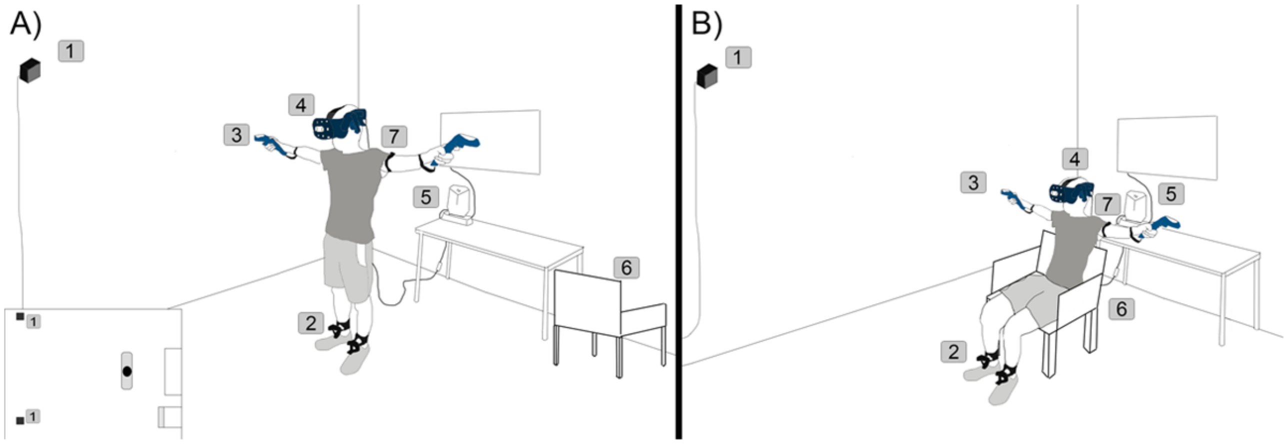
We measured fear of falling by asking participants "how concerned are you about the possibility of falling during the experiment?" using a 10-point Likert scale from 1 to 10, with 1 indicating "very slightly or not at all" and 10 indicating "extremely."

After completing the above questionnaires, participants were asked to answer the following open-ended question in the questionnaire: "What do you think about this version of the game?" They responded by typing into a text box. There was no limit for the length of participants' responses. A full list of questions used after each condition can be found in Multimedia Appendix 2.

Apparatus and Setup

The experiment was conducted using HTC VIVE Pro Eye connected to an HP Z workstation (i7 CPU, 16 GB RAM, and Nvidia Quadro P5200 GPU). Two HTC VIVE handheld controllers, two HTC VIVE trackers, and two base stations were used to enable hand and feet motion tracking. A stable chair with two handles was used in the sitting condition. The HR was monitored by a Polar OH1 optical HR sensor, which has been proved to be able to capture good HR data when compared with the gold standard of HR measurement with an electrocardiography device [62,63]. The experiment was conducted in an indoor laboratory room that could not be seen from the outside. The room temperature was always set to be 23 to 24°C during the experiment. Figure 3 depicts the experiment setup and devices used in the experiment.

Figure 3. Standing (A) and seated (B) experiment setup, and the devices used in the experiment: (1) the HTC VIVE base stations, the locations of the two base stations can be found in the vertical view in the left bottom corner; (2) the two VIVE trackers attached to the legs; (3) the VIVE controllers used to track the player's hand positions; (4) HTC VIVE Pro Eye; (5) the HP Z backpack; (6) the chair; and (7) Polar OH1.



Procedure

Before filling in the pre-experiment questionnaire that gathered demographic information (eg, age, gender, and experience with the VR device), we obtained participants' consent to participate in the experiment and collected their RestHR and resting blood pressure. They were also asked to enter their age, gender, height, and weight into the Polar Beat app. Before each condition started, a researcher would help them to wear the VIVE Pro Eye headset with two VIVE handheld controllers and two VIVE trackers. Participants were then given 3 minutes to get familiar with the corresponding condition of the iVR exergame. Once their HR reached the equivalent RestHR level, they proceeded to the experiment stage, beginning with calibrating the game, training, and testing. In each condition, they needed to perform 120 gestures, requiring 5 minutes (120 gestures × 2.5 seconds). We fixed the number of gestures to allow for comparing the two games. After each condition, they were asked to fill in postexperiment questionnaires. They proceeded to the next condition when they felt rested and their HR was at the rest level.

Statistical Analysis

We used the paired *t* test to understand the difference between the seated exergame and standing exergame regarding gameplay performance, IMI, MSAQ, fear of falling, and exertion measurements. For the percentage of missed gesture types for sitting/standing gesture type, we used one-way repeated analysis of variance (ANOVA) with the sitting/standing gesture type as the within-subjects variables, respectively. We also examined and reported if there were any significant gender differences in our measurements by using one-way between-subjects ANOVA. We set the α level at .05 in our analyses. We further reported the effect sizes using Cohen suggestion to classify the effect size, where Cohen suggested that d=0.2 represents a "small" effect size, 0.5 represents a "medium" effect size, and 0.8 represents a "large" effect size [64]. Analyses were performed using the Statistical Package for the Social Sciences (IBM Corp).

Results

Participant Characteristics

Fifty-four individuals were interested in participating in the experiment. Two were excluded owing to their high RestHR. At the end, a total of 52 participants were eligible to participate in the study. The characteristics of the study participants are shown in Table 2.

Table 2. Characteristics of the study participants.

Characteristic	Value
Number of students, n	52
Age (years), mean (SD)	18.81 (1.70)
RestHR[a], mean (SD)	77.71 (8.78)
Height (cm), mean (SD)	170.11 (7.75)
Weight (kg), mean (SD)	60.84 (10.79)
Body mass index (kg/m^2), mean (SD)	20.93 (2.87)
Self-reported exercise time per week (min), mean (SD)	87.88 (66.55)
Normal or corrected-to-normal, n	52
Self-reported experience with seated exercise regimes	No
Self-reported experience with VR[b] HMDs[c], n	21
Frequent user, n	1
Self-reported experience with full-body gesture-based video games, n	28
Frequent player, n	4

[a]RestHR: resting heart rate.

[b]VR: virtual reality.

[c]HMD: head-mounted display.

Gameplay Data

Gameplay data and analysis results are reported in Table 3. The analysis showed that game mode did not influence action completion time. However, the analysis showed that players missed more gestures in the seated exergame than in the standing exergame.

Table 3. Gameplay data and exertion measures.

Variable	Standing exergame, mean (SD)	Seated exergame, mean (SD)	t_{51}	*P* value[a]	Cohen *d*
Gameplay					
Action completion time	1.48 (0.14)	1.53 (0.18)	1.87	.07	N/A[b]
Missed gestures	1.65 (1.71)	2.33 (1.83)	2.40	.02	0.332
Exertion					
AvgHR%[c]	51.9% (4.6%)	54.3% (5.0%)	4.66	<.001	0.646
Increased HR%[d]	6.9% (4.4%)	11.8% (5.3%)	5.86	<.001	0.813
Calories	21.83 (6.76)	24.67 (7.25)	4.44	<.001	0.615
Borg 6-20	9.02 (2.15)	10.25 (2.59)	3.96	<.001	0.548

[a]Significant at .05.

[b]N/A: not applicable.

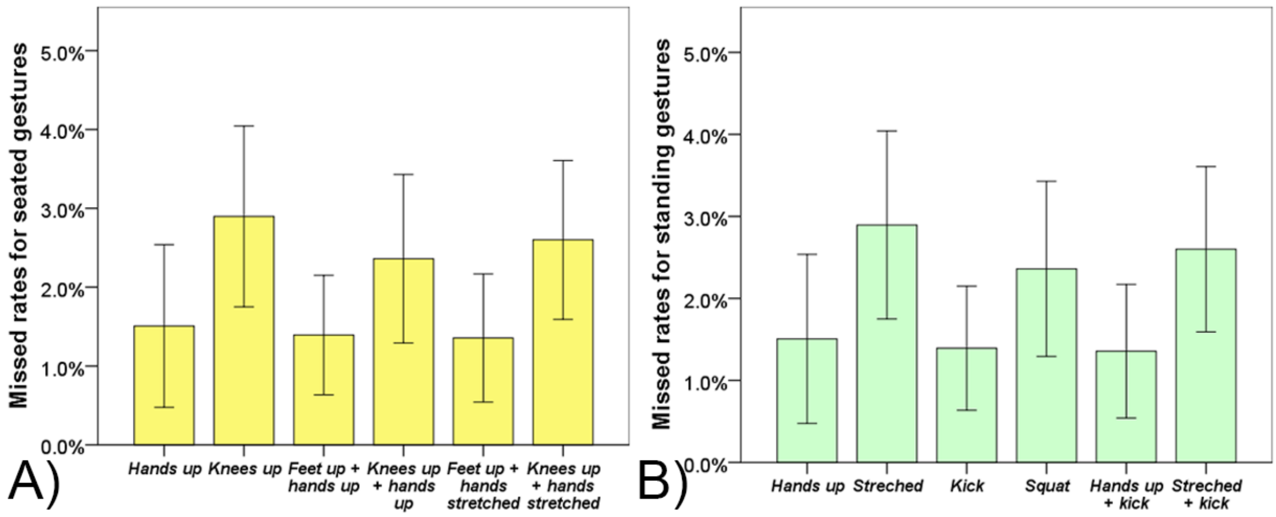
[c]AvgHR%: average heart rate percentage.

[d]HR%: heart rate percentage.

Percentage of Missed Gesture Types

The results of one-way repeated ANOVA yielded no significant effect of the sitting gesture type ($F_{5,255}$=1.98, P=.08) or standing gesture type ($F_{5,255}$=1.058, P=.38) on the percentage of corresponding missed gestures. The missed rate for sitting and standing gesture types can be found in Figure 4.

Figure 4. Missed rates for (A) seated gestures and (B) standing gestures. Error bars indicate ±2 standard errors.

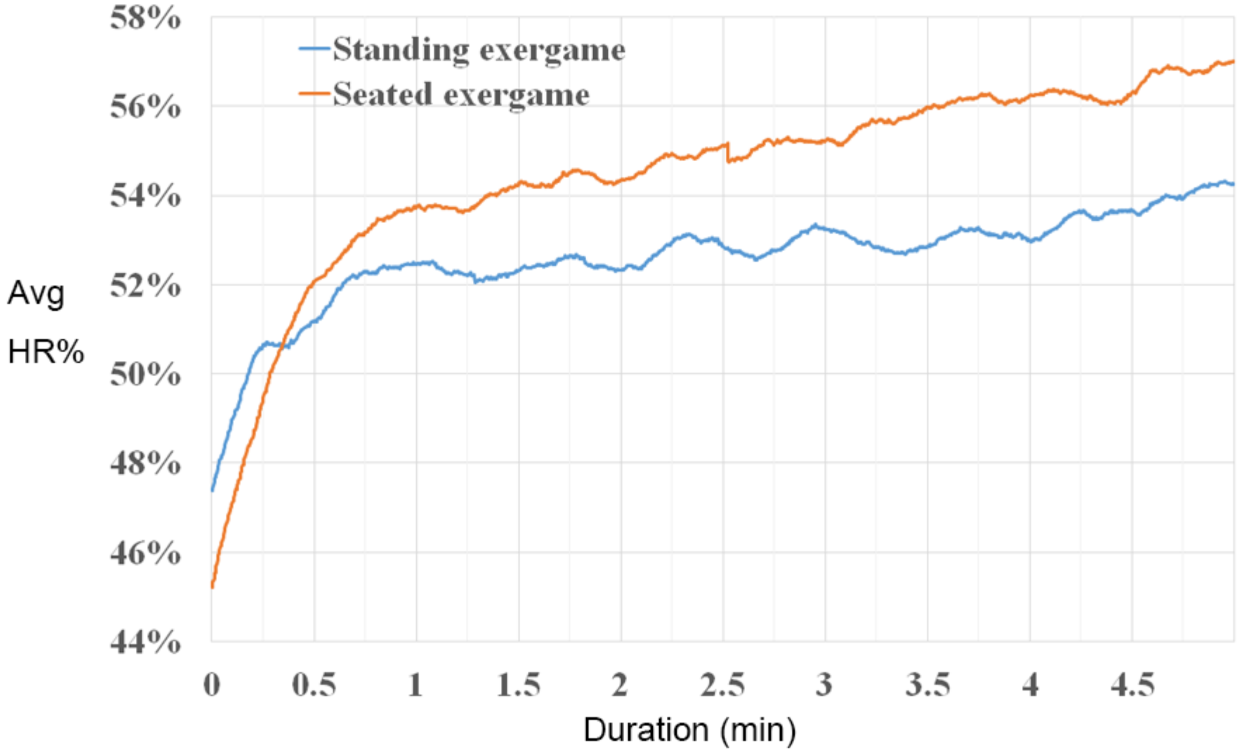


Exertion

Exertion (AvgHR%, increased HR%, calories burned, and Borg 6-20) data and analysis results are presented in Table 3. The analysis yielded significantly higher AvgHR%, increased HR%, calories burned, and Borg 6-20 RPE for the seated exergame (all $P<.001$). Our results suggest that the seated exergame was rated as "very light" to "light" exercise and the standing exergame was rated as "very light" exercise according to the Borg 6-20 RPE scale.

To aid the visualization of the AvgHR% behavior of both exergames, Figure 5 shows the AvgHR% data from all participants during the 5 minutes of gameplay, averaged over the whole session. The seated exergame had a higher AvgHR% than the standing exergame after 0.34 minutes.

Figure 5. Mean AvgHR% during gameplay for both versions of the exergame; the interaction between two lines occurs at 0.38 minutes. At 2.64 minutes, AvgHR% of the seated exergame reached the moderate physical intensity level. AvgHR%: average heart rate percentage.



Experience

Data analysis of the MSAQ included the overall MSAQ score and its subscale scores (gastrointestinal, central, peripheral, and sopite-related). The MSAQ data and analysis results are reported in Table 4. The analysis showed significantly higher peripheral sickness (P=.02) and sopite-related sickness (P=.004) for the seated exergame. We did not find a significant difference between the seated exergame and standing exergame in terms of central (P=.81), gastrointestinal (P=.81), and overall sickness (P=.06).

Regarding IMI, there was no significant difference for interest/enjoyment (P=.96), perceived competence (P=.26), and pressure/tension (P=.42). However, the analysis yielded a significantly higher value/usefulness for the seated exergame (P=.04). IMI data and analysis results can be found in Table 4.

The one-way between-subjects ANOVA showed that interest/enjoyment was rated significantly higher by females (mean 5.20, SD 1.23) than males (mean 4.38, SD 41.16), with a medium effect size ($F_{1,50}$=6.01, P=.02).

Table 4. Motion Sickness Assessment Questionnaire and Intrinsic Motivation Inventory test measures.

Variable	Standing exergame, mean (SD)	Sitting exergame, mean (SD)	t_{51}	P value[a]	Cohen d
MSAQ[b]					
Peripheral	22.7% (12.9%)	27.7% (14.3%)	−2.41	.02[a]	0.334
Sopite-related	18.8% (11.3%)	23.9% (15.4%)	−3.06	.004[a]	0.424
Central	19.2% (11.1%)	18.8% (10.7%)	0.24	.81	N/A[c]
Gastrointestinal	13.5% (5.7%)	13.7% (5.2%)	−0.24	.81	N/A
Overall	17.4% (7.3%)	19.4% (8.4%)	−1.91	.061	N/A
IMI[d]					
Interest/enjoyment	4.73 (1.30)	4.72 (1.30)	0.05	.96	N/A
Competence	4.99 (1.12)	4.85 (1.20)	1.14	.26	N/A
Pressure/tension	2.90 (0.95)	2.99 (1.04)	−0.82	.42	N/A
Value/usefulness	5.12 (1.28)	5.38 (1.12)	−2.11	.04[a]	0.292

[a]Significant at .05.

[b]MSAQ: Motion Sickness Assessment Questionnaire.

[c]N/A: not applicable.

[d]IMI: Intrinsic Motivation Inventory.

Fear of Falling

There was no significant difference in the fear of falling ratings between the standing exergame (mean 2.10, SD 1.58) and seated exergame (mean 2.40, SD 1.78) (t_{51}=−1.16, P=.25).

Discussion

Overview

With the limited exploration of seated exergames in the literature of iVR exergames, this study is the first to explore the difference between full-body gesture-based seated exergames and standing exergames in iVR among university students regarding playability (ie, gameplay performance) and user experience (ie, intrinsic motivation and motion sickness). Our results suggest that participants perceived higher value in the seated exergame than in the standing exergame. However, the seated exergame was associated with a worse gameplay performance (ie, the number of missed gestures) and a higher rating of motion sickness than the standing exergame.

Although we observed that participants missed a significantly higher number of gestures in the seated exergame than in the standing exergame (P=.02), this rate was as low as 1.9% (2.3/120). Further analysis of the type of gestures missed in the seated exergame confirmed that these misses were in the early stages of the experiment, and as such, the main reason for the misses could be because of participant unfamiliarity with exercising in the seated position (none of them had previous experience of seated exercising; Table 2).

Regarding motion sickness, previous studies [31,32] have suggested that the seated exergame might result in a lower level of motion sickness. However, this was not supported by our findings. We observed that participants felt sicker (ie, peripheral and sopite-related motion sickness) in the seated exergame. However, the reason was beyond the scope of this study; a further investigation is required to understand why motion sickness was higher in the seated exergame than in the standing exergame. We suggest that future designers and researchers should carefully design full-body gestures for iVR seated exergames to minimize motion sickness.

As for intrinsic motivation, we did not observe any significant difference between the seated exergame and standing exergame (ie, interest/enjoyment, P=.96; perceived competence, P=.26; pressure/tension, P=.42). However, there was a gender effect on participants' intrinsic motivation toward exergames, where we found that females had a significantly higher intrinsic motivation (ie, interest/enjoyment) than males (P=.02). This could be because the exergame involved in our study was more like a dance game, which was inspired by *Just Dance*, and prior research [65] has shown that females tend to be more physically active in dance exergames. Aside from this difference between male and female participants, no other differences were found in our experiment.

In most cases, standing exercises have a higher exercise intensity (in traditional exercises [66,67] and exergames [68,69]). We found that our seated exergame led to a higher exertion (ie, AvgHR%, increased HR%, calories burned, and Borg 6-20) than the standing exergame, possibly because the seated exergame involved more whole-body movements that required increased energy expenditure during gameplay [70-73].

Design Guidelines

In this section, we provide design guidelines that are based on suggestions provided by Wiemeyer et al [74] for future game designers who are interested in building iVR full-body standing or seated gesture-based exergames.

First, practice should be provided for each gesture during warm-up. A warm-up session before exercising is essential [75], and it should be included in exergames as well. One way to perform warm-up for full-body gesture-based exergames is to practice the gestures involved in the game, which will not only help players reduce the risk of injures but also make them familiar with the in-game gestures.

Second, the difficulty level of the game should be adapted to the current state of the individual. Regarding an *offline* approach, players might have difficulty in performing certain gestures during gameplay. Therefore, to match the difficulty of the game to the current state of the individual, it would be necessary for players to experience and select gestures they are comfortable performing before playing the game. Regarding an *online* approach, one of the adaptive methods that has been used and proven to be suitable in exergames is proportional–integral–derivative (PID) control [76]. Designers can use PID control to modify the transition time between gestures or select the gestures based on the player's real-time HR and gameplay performance (ie, the number of gestures missed). PID control is also useful to avoid overly vigorous exercise, which might put the exerciser at risk of eliciting unwanted coronary issues [77].

Third, warning signs should be provided for standing exergames if users have left (or are about to leave) the calibration position and are too far to keep them protected. This is because players tend to move around during gameplay, which we encountered in our study and has been reported previously [23]. It could lead to potentially dangerous situations (eg, hitting objects that are in the environment and going out of the safe tracking area) or decrease the recognition performance of the sensors (eg, tracking may not work when they are too close to or far from the sensors).

Limitations and Future Work

Our study only focused on one sedentary lifestyle user group (university students). Future work could focus on investigating the two versions of the exergame with different population groups (eg, older adults and users who have physical disabilities). To minimize players' cognitive workload, we set both exergames to include only 6 out of the 10 gesture types that we measured during the pilot study. In the future, we could add more gestures to increase the complexity of the game (as stated by participants 13, 20, 30, and 37). A further limitation is that our experiment did not measure which types of gestures caused the unwanted level of motion sickness in the seated exergame. Future experiments could be conducted to check issues related to motion sickness based on specific gestures and types of gestures.

Conclusions

Our contributions to the field of iVR exergaming regarding gameplay performance and user experience are as follows: (1) the iVR seated exergame could result in higher exertion and provide higher value to players than the standing exergame; (2) participants might feel sicker in the iVR seated exergame than in the standing exergame, and as such, full-body gestures for seated exergames need to be designed carefully to help minimize the feeling of motion sickness; and (3) participants might miss more gestures in the iVR seated exergame than in the standing exergame. Therefore, designers should allow more time for performing gestures in the seated exergame.

Acknowledgments

This research was supported by Xi'an Jiaotong-Liverpool University Key Program Special Fund (grant no. KSF-A-03) and Xi'an Jiaotong-Liverpool University Research Development Fund. The authors thank all the participants who volunteered for the study and the reviewers for their suggestions that helped improve this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Approach for measuring the intensity of gestures.
[PDF File (Adobe PDF File), 122 KB - games_v8i3e17972_app1.pdf]

Multimedia Appendix 2
Questionnaire used in the study.
[PDF File (Adobe PDF File), 86 KB - games_v8i3e17972_app2.pdf]

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet 2012 Jul;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9]
2. Williams G, Fruhbeck G. Obesity: Science to Practice. West Sussex, UK: Wiley-Blackwell; 2009:607.
3. Hernandez H, Ye Z, Graham T, Fehlings D, Switzer L. Designing action-based exergames for children with cerebral palsy. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 1261-1270. [doi: 10.1145/2470654.2466164]
4. Xu W, Liang H, Zhang Z, Baghaei N. Studying the Effect of Display Type and Viewing Perspective on User Experience in Virtual Reality Exergames. Games Health J 2020 Feb 19. [doi: 10.1089/g4h.2019.0102] [Medline: 32074463]
5. Gerling K, Livingston I, Nacke L, Mandryk R. Full-body motion-based game interaction for older adults. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, Texas, USA p. 1873-1882. [doi: 10.1145/2207676.2208324]
6. Bogost I. The Rhetoric of Exergaming. In: Proceedings of the Digital Arts and Cultures. 2005 Presented at: Digital Arts and Cultures; December 1-3, 2005; Copenhagen, Denmark p. 10.
7. Sinclair J, Hingston P, Masek M. Considerations for the design of exergames. In: Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia. 2007 Presented at: 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia; December 1-4, 2007; Perth, Australia p. 289-295. [doi: 10.1145/1321261.1321313]
8. Farrow M, Lutteroth C, Rouse PC, Bilzon JL. Virtual-reality exergaming improves performance during high-intensity interval training. Eur J Sport Sci 2019 Jul;19(6):719-727. [doi: 10.1080/17461391.2018.1542459] [Medline: 30403927]
9. Moholdt T, Weie S, Chorianopoulos K, Wang AI, Hagen K. Exergaming can be an innovative way of enjoyable high-intensity interval training. BMJ Open Sport Exerc Med 2017;3(1):e000258 [FREE Full text] [doi: 10.1136/bmjsem-2017-000258] [Medline: 29021913]
10. Pluchino A, Lee SY, Asfour S, Roos BA, Signorile JF. Pilot study comparing changes in postural control after training using a video game balance board program and 2 standard activity-based balance intervention programs. Arch Phys Med Rehabil 2012 Jul;93(7):1138-1146. [doi: 10.1016/j.apmr.2012.01.023] [Medline: 22414490]
11. Rosenberg D, Depp CA, Vahia IV, Reichstadt J, Palmer BW, Kerr J, et al. Exergames for Subsyndromal Depression in Older Adults: A Pilot Study of a Novel Intervention. The American Journal of Geriatric Psychiatry 2010 Mar;18(3):221-226. [doi: 10.1097/jgp.0b013e3181c534b5]
12. Zeng N, Pope Z, Lee J, Gao Z. Virtual Reality Exercise for Anxiety and Depression: A Preliminary Review of Current Research in an Emerging Field. J Clin Med 2018 Mar 04;7(3) [FREE Full text] [doi: 10.3390/jcm7030042] [Medline: 29510528]
13. Gao Y, Mandryk RL. GrabApple: The Design of a Casual Exergame. In: Anacleto JC, Fels S, Graham N, Kapralos B, Saif El-Nasr M, Stanley K, editors. Entertainment Computing – ICEC 2011. Heidelberg, Berlin: Springer; 2011.
14. Faiola A, Kharrazi H. Diabetes education and serious gaming: Teaching adolescents to cope with diabetes. In: Health Informatics: A Patient-Centered Approach to Diabetes. Cambridge, Massachusetts: MIT Press; 2010:151-177.
15. Peng W, Lin J, Crouse J. Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):681-688. [doi: 10.1089/cyber.2010.0578] [Medline: 21668370]
16. Huang H, Nguyen HV, Cheng T, Wong M, Chiu H, Yang Y, et al. A Randomized Controlled Trial on the Role of Enthusiasm About Exergames: Players' Perceptions of Exercise. Games Health J 2019 Jun;8(3):220-226. [doi: 10.1089/g4h.2018.0057] [Medline: 30418041]
17. Sápi M, Domján A, Fehérné Kiss A, Pintér S. Is Kinect Training Superior to Conventional Balance Training for Healthy Older Adults to Improve Postural Control? Games Health J 2019 Feb;8(1):41-48. [doi: 10.1089/g4h.2018.0027] [Medline: 30153062]
18. da Silva Alves R, Iunes DH, de Carvalho JM, Menezes FD, Silva AM, Borges JB, et al. Effects of Exergaming on Quality of Life in Cancer Patients. Games for Health Journal 2018 Dec;7(6):385-392. [doi: 10.1089/g4h.2017.0174]
19. Garcia JA, Schoene D, Lord SR, Delbaere K, Valenzuela T, Navarro KF. A Bespoke Kinect Stepping Exergame for Improving Physical and Cognitive Function in Older People: A Pilot Study. Games Health J 2016 Dec;5(6):382-388. [doi: 10.1089/g4h.2016.0070] [Medline: 27860515]
20. Xie B, Zhang Y, Huang H, Ogawa E, You T, Yu L. Exercise Intensity-Driven Level Design. IEEE Trans. Visual. Comput. Graphics 2018 Apr;24(4):1661-1670. [doi: 10.1109/tvcg.2018.2793618]
21. Xu W, Liang H, Zhao Y, Yu D, Monteiro D. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2019 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-14. [doi: 10.1145/3290605.3300674]
22. Xu W, Liang H, Yu Y, Monteiro D, Hasan K, Fleming C. Assessing the Effects of a Full-body Motion-based Exergame in Virtual Reality. In: Proceedings of the Seventh International Symposium of Chinese CHI. 2019 Presented at: Seventh International Symposium of Chinese CHI; June 27-30, 2019; Xiamen, China p. 1-6. [doi: 10.1145/3332169.3333574]

23. Ioannou C, Archard P, O?Neill E, Lutteroth C. Virtual Performance Augmentation in an Immersive Jump & Run Exergame. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2019 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-15. [doi: 10.1145/3290605.3300388]
24. Granqvist A, Takala T, Takatalo J, Hämäläinen P. Exaggeration of Avatar Flexibility in Virtual Reality. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2018 Presented at: Annual Symposium on Computer-Human Interaction in Play; October 28-31, 2018; Melbourne, Australia p. 201-209. [doi: 10.1145/3242671.3242694]
25. Shaw LA, Marks S, Wunsche BC, Lutteroth C, Buckley J, Corballis P. Development and Evaluation of an Exercycle Game Using Immersive Technologies. In: Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management. 2015 Jan Presented at: The 8th Australasian Workshop on Health Informatics and Knowledge Management; 2015; Sydney, Australia p. 11.
26. Alexandris K, Tsorbatzoudis C, Grouios G. Perceived Constraints on Recreational Sport Participation: Investigating their Relationship with Intrinsic Motivation, Extrinsic Motivation and Amotivation. Journal of Leisure Research 2017 Dec 13;34(3):233-252. [doi: 10.1080/00222216.2002.11949970]
27. Frederick C, Ryan R. Differences in Motivation for Sport and Exercise and their Relations with Participation and Mental Health. Journal of Sport Behavior 1993 Sep 1;16(3):124-146 [FREE Full text]
28. Ryan R, Frederick C, Lepes D, Rubio N, Sheldon K. Intrinsic Motivation and Exercise Adherence. International Journal of Sport Psychology 1997 Oct;28(4):335-354 [FREE Full text]
29. Gómez-López M, Gallegos AG, Extremera AB. Perceived barriers by university students in the practice of physical activities. J Sports Sci Med 2010;9(3):374-381 [FREE Full text] [Medline: 24149629]
30. Gerling K, Hicks K, Kalyn M, Evans A, Linehan C. Designing Movement-based Play With Young People Using Powered Wheelchairs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2016 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2016; New York, NY, USA p. 4447-4458. [doi: 10.1145/2858036.2858070]
31. Bolton J, Lambert M, Lirette D, Unsworth B. PaperDude: a virtual reality cycling exergame. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems-Extended Abstracts. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Ontario, Canada p. 475-478. [doi: 10.1145/2559206.2574827]
32. Shaw L, Wunsche B, Lutteroth C, Marks S, Callies R. Challenges in Virtual Reality Exergame Design. In: Proceedings of the Australasian User Interface Conference. 2015 Presented at: Australasian User Interface Conference; January 27-30, 2015; Sydney, Australia p. 61-68.
33. Ryan RM, Deci EL. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
34. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. New York, USA: Springer; 1985:372.
35. Reeve J, Deci EL. Elements of the Competitive Situation that Affect Intrinsic Motivation. Pers Soc Psychol Bull 1996;22(1):24-33. [doi: 10.1177/0146167296221003]
36. Finkelstein S, Suma EA. Astrojumper: Motivating Exercise with an Immersive Virtual Reality Exergame. Presence: Teleoperators and Virtual Environments 2011 Feb;20(1):78-92. [doi: 10.1162/pres_a_00036]
37. Bailey BW, McInnis K. Energy cost of exergaming: a comparison of the energy cost of 6 forms of exergaming. Arch Pediatr Adolesc Med 2011 Jul;165(7):597-602. [doi: 10.1001/archpediatrics.2011.15] [Medline: 21383255]
38. Baños RM, Escobar P, Cebolla A, Guixeres J, Alvarez Pitti J, Lisón JF, et al. Using Virtual Reality to Distract Overweight Children from Bodily Sensations During Exercise. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):115-119. [doi: 10.1089/cyber.2015.0283] [Medline: 26882326]
39. Monedero J, Lyons EJ, O'Gorman DJ. Interactive video game cycling leads to higher energy expenditure and is more enjoyable than conventional exercise in adults. PLoS One 2015;10(3):e0118470 [FREE Full text] [doi: 10.1371/journal.pone.0118470] [Medline: 25738290]
40. Irwin JD. The prevalence of physical activity maintenance in a sample of university students: a longitudinal study. J Am Coll Health 2007;56(1):37-41. [doi: 10.3200/JACH.56.1.37-42] [Medline: 17711824]
41. Xu W, Liang H, Baghaei N, Wu Berberich B, Yue Y. Health Benefits of Digital Videogames for the Aging Population: A Systematic Review. Games Health J 2020 Jun 24. [doi: 10.1089/g4h.2019.0130] [Medline: 32589482]
42. Iskenderova A, Weidner F, Broll W. Drunk Virtual Reality Gamingxploring the Influence of Alcohol on Cybersickness. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: Annual Symposium on Computer-Human Interaction in Play; November 15-18, 2017; Amsterdam Netherlands p. 561-572. [doi: 10.1145/3116595.3116618]
43. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
44. Ostchega Y, Porter KS, Hughes J, Dillon CF, Nwankwo T. Resting pulse rate reference data for children, adolescents, and adults: United States, 1999-2008. Natl Health Stat Report 2011 Aug 24(41):1-16 [FREE Full text] [Medline: 21905522]
45. Rogers S. Level Up! The Guide to Great Video Game Design. West Sussex, UK: John Wiley & Sons; Apr 2014:552.

46. Lin J. "Just Dance": The Effects of Exergame Feedback and Controller Use on Physical Activity and Psychological Outcomes. Games Health J 2015 Jun;4(3):183-189. [doi: 10.1089/g4h.2014.0092] [Medline: 26182062]
47. Xu W, Liang H, Zhao Y, Zhang T, Yu D, Monteiro D, et al. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. IEEE Trans Vis Comput Graph 2019 May 1;25(5):1991-2001. [doi: 10.1109/TVCG.2019.2898736] [Medline: 30762551]
48. Farzan R, DiMicco J, Millen D, Dugan C, Geyer W, Brownholtz E. Results from Deploying a Participation Incentive Mechanism within the Enterprise. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 5-10, 2008; Florence, Italy p. 563-572. [doi: 10.1145/1357054.1357145]
49. Kluwer W. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, US: Lippincott Williams & Wilkins; Feb 9, 2013:480.
50. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
51. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
52. Gianaros PJ, Muth ER, Mordkoff JT, Levine ME, Stern RM. A questionnaire for the assessment of the multiple dimensions of motion sickness. Aviat Space Environ Med 2001 Feb;72(2):115-119 [FREE Full text] [Medline: 11211039]
53. Speicher M, Feit A, Ziegler P, Krüger A. Selection-based Text Entry in Virtual Reality. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2018 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC, Canada p. 1-13. [doi: 10.1145/3173574.3174221]
54. Xu W, Liang H, He A, Wang Z. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In: Proceeding of IEEE International Symposium on Mixed and Augmentation Reality. 2019 Presented at: IEEE International Symposium on Mixed and Augmentation Reality; October 14-18, 2019; Beijing, China p. 11. [doi: 10.1109/ismar.2019.00026]
55. Frederiksen JG, Sørensen SM, Konge L, Svendsen MB, Nobel-Jørgensen M, Bjerrum F, et al. Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. Surg Endosc 2020 Mar;34(3):1244-1252. [doi: 10.1007/s00464-019-06887-8] [Medline: 31172325]
56. Ryan RM. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology 1982;43(3):450-461. [doi: 10.1037/0022-3514.43.3.450]
57. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [doi: 10.1111/j.1467-6494.1994.tb00797.x] [Medline: 8169757]
58. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
59. Clancy RB, Herring MP, Campbell MJ. Motivation Measures in Sport: A Critical Review and Bibliometric Analysis. Front Psychol 2017;8:348 [FREE Full text] [doi: 10.3389/fpsyg.2017.00348] [Medline: 28337165]
60. Barathi S, O?Neill E, Lutteroth C, Finnegan D, Farrow M, Whaley A, et al. Interactive Feedforward for Improving Performance and Maintaining Intrinsic Motivation in VR Exergaming. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2018 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC, Canada p. 1-14. [doi: 10.1145/3173574.3173982]
61. Keesing A, Ooi M, Wu O, Ye X, Shaw L, Wünsche B. HIIT With Hits: Using Music and Gameplay to Induce HIIT in Exergames. In: Proceedings of the Australasian Computer Science Week Multiconference. 2019 Presented at: Australasian Computer Science Week Multiconference; January 2019; Sydney, Australia p. 1-10.
62. Hettiarachchi IT, Hanoun S, Nahavandi D, Nahavandi S. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. PLoS One 2019;14(5):e0217288 [FREE Full text] [doi: 10.1371/journal.pone.0217288] [Medline: 31120968]
63. Schubert M, Clark A, De La Rosa A. The Polar OH1 Optical Heart Rate Sensor is Valid during Moderate-Vigorous Exercise. Sports Med Int Open 2018 Jun;2(3):E67-E70 [FREE Full text] [doi: 10.1055/a-0631-0920] [Medline: 30539120]
64. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Abingdon, UK: Routledge; Jul 1, 1988:400.
65. Gao Z, Podlog L, Lee J. Children's situational motivation, rate of perceived exertion and physical activity levels in exergaming: Associations and gender differences. In: Video Games: Parents' Perceptions, Role of Social Media and Effects on Behavior. Hauppauge, NY, USA: Nova Science Publishers, Inc; 2014:17-28.
66. Montoye H. Energy costs of exercise and sport. In: Nutrition in Sport. Oxford, UK: Blackwell; Jan 1, 2000:53-72.
67. Reiff C, Marlatt K, Dengel DR. Difference in caloric expenditure in sitting versus standing desks. J Phys Act Health 2012 Sep;9(7):1009-1011. [doi: 10.1123/jpah.9.7.1009] [Medline: 22971879]
68. Malone LA, Thirumalai M, Padalabalanarayanan S, Neal WN, Bowman S, Mehta T. Energy Expenditure and Enjoyment During Active Video Gaming Using an Adapted Wii Fit Balance Board in Adults with Physical Disabilities: Observational Study. JMIR Serious Games 2019 Feb 01;7(1):e11326 [FREE Full text] [doi: 10.2196/11326] [Medline: 30707098]
69. Mat Rosly M, Mat Rosly H, Davis Oam GM, Husain R, Hasnan N. Exergaming for individuals with neurological disability: a systematic review. Disabil Rehabil 2017 Apr;39(8):727-735. [doi: 10.3109/09638288.2016.1161086] [Medline: 27108475]

70. Graves LE, Ridgers ND, Stratton G. The contribution of upper limb and total body movement to adolescents' energy expenditure whilst playing Nintendo Wii. Eur J Appl Physiol 2008 Nov;104(4):617-623. [doi: 10.1007/s00421-008-0813-8] [Medline: 18607619]
71. Lanningham-Foster L, Foster RC, McCrady SK, Jensen TB, Mitre N, Levine JA. Activity-promoting video games and increased energy expenditure. J Pediatr 2009 Jun;154(6):819-823 [FREE Full text] [doi: 10.1016/j.jpeds.2009.01.009] [Medline: 19324368]
72. Lanningham-Foster L, Jensen TB, Foster RC, Redmond AB, Walker BA, Heinz D, et al. Energy expenditure of sedentary screen time compared with active screen time for children. Pediatrics 2006 Dec;118(6):e1831-e1835. [doi: 10.1542/peds.2006-1087] [Medline: 17142504]
73. Mellecker RR, McManus AM. Energy expenditure and cardiovascular responses to seated and active gaming in children. Arch Pediatr Adolesc Med 2008 Sep;162(9):886-891. [doi: 10.1001/archpedi.162.9.886] [Medline: 18762609]
74. Wiemeyer J, Deutsch J, Malone LA, Rowland JL, Swartz MC, Xiong J, et al. Recommendations for the Optimal Design of Exergame Interventions for Persons with Disabilities: Challenges, Best Practices, and Future Research. Games Health J 2015 Feb;4(1):58-62 [FREE Full text] [doi: 10.1089/g4h.2014.0078] [Medline: 26181682]
75. Bushman B. ACSM's Complete Guide to Fitness & Health. Champaign, USA: Human Kinetics; Feb 9, 2017:448.
76. Muñoz J, Cameirão M, Bermúdez i Badia S, Gouveia E. Closing the Loop in Exergaming - Health Benefits of Biocybernetic Adaptation in Senior Adults. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2018 Presented at: Annual Symposium on Computer-Human Interaction in Play; October 28-31, 2018; Melbourne, Australia p. 329-339. [doi: 10.1145/3242671.3242673]
77. Möhlenkamp S, Lehmann N, Breuckmann F, Bröcker-Preuss M, Nassenstein K, Halle M, Marathon Study Investigators, Heinz Nixdorf Recall Study Investigators. Running: the risk of coronary events : Prevalence and prognostic relevance of coronary atherosclerosis in marathon runners. Eur Heart J 2008 Aug;29(15):1903-1910. [doi: 10.1093/eurheartj/ehn163] [Medline: 18426850]

Abbreviations

ANOVA: analysis of variance
AvgHR%: average heart rate percentage
HMD: head-mounted display
HR: heart rate
IMI: Intrinsic Motivation Inventory
iVR: immersive virtual reality
MaxHR: maximum heart rate
MSAQ: Motion Sickness Assessment Questionnaire
PAR-Q: Physical Activity Readiness Questionnaire
PID: proportional–integral–derivative
RestHR: resting heart rate
RPE: rating of perceived exertion
VR: virtual reality

Edited by G Eysenbach; submitted 25.01.20; peer-reviewed by A Duhamel, E Dove; comments to author 16.03.20; revised version received 27.04.20; accepted 03.06.20; published 27.07.20.

Please cite as:
Xu W, Liang HN, He Q, Li X, Yu K, Chen Y
Results and Guidelines From a Repeated-Measures Design Experiment Comparing Standing and Seated Full-Body Gesture-Based Immersive Virtual Reality Exergames: Within-Subjects Evaluation
JMIR Serious Games 2020;8(3):e17972
URL: http://games.jmir.org/2020/3/e17972/
doi:10.2196/17972
PMID:32716004

Original Paper

Parents of Adolescents Perspectives of Physical Activity, Gaming and Virtual Reality: Qualitative Study

Lucy McMichael[1*], BSc, MSc; Nuša Farič[1*], BSc, MSc; Katie Newby[2*], BSc, MSc, PhD; Henry W W Potts[3*], BA, MSc, PhD; Adrian Hon[4*], BSc, MSc; Lee Smith[5*], BSc, MSc, PhD; Andrew Steptoe[1*], BA, MA, DPhil, DSC; Abi Fisher[1*], BSc, PhD

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom
[2]Department of Psychology and Sport Sciences, University of Hertfordshire, Hertfordshire, United Kingdom
[3]Institute of Health Informatics, University College London, London, United Kingdom
[4]Six to Start, Redditch, United Kingdom
[5]The Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Nuša Farič, BSc, MSc
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London
United Kingdom
Phone: 44 020776794466 ext 41914
Email: nfaric@gmail.com

Abstract

Background: Virtual reality (VR) exergaming may be a promising avenue to engage adolescents with physical activity. Since parental support is a consistent determinant of physical activity in adolescents, it is crucial to gather the views of parents of adolescents about this type of intervention.

Objective: This study aimed to interview parents of younger adolescents (13-17 years old) about physical activity, gaming, and VR as part of the larger vEngage study.

Methods: Semistructured interviews were conducted with 18 parents of adolescents. Data were synthesized using framework analysis.

Results: Parents believed that encouraging physical activity in adolescents was important, particularly for mental health. Most parents felt that their children were not active enough. Parents reported their adolescents regularly gamed, with mostly negative perceptions of gaming due to violent content and becoming addicted. Parents discussed an inability to relate to gaming due to "generational differences," but an exception was exergaming, which they had played with their children in the past (eg, Wii Fit). Specific recommendations for promoting a VR exergaming intervention were provided, but ultimately parents strongly supported harnessing gaming for any positive purpose.

Conclusions: The current study suggests promise for a VR exergaming intervention, but this must be framed in a way that addresses parental concerns, particularly around addiction, violence, and safety, without actively involving their participation. While parents would rather their children performed "real-world" physical activity, they believed the key to engagement was through technology. Overall, there was the perception that harnessing gaming and sedentary screen time for a positive purpose would be strongly supported.

(JMIR Serious Games 2020;8(3):e14920) doi:10.2196/14920

KEYWORDS

exercise; obesity; video games; adolescent; adolescence; sports; health; leisure activities; virtual reality

Introduction

Sufficient physical activity protects against noncommunicable diseases [1] and is associated with better mental health [2]. The detrimental health outcomes associated with low population levels of physical activity have placed a significant strain on the economy and health services [3]. Adolescence, which spans the ages of 10-24 years, encompasses an important phase of social and biological development and has been suggested as the stage at which individuals acquire skills that contribute to future health and wellbeing [4]. Thus, adolescence is an opportune time to encourage engagement with physical activity [5]. Early adolescence is particularly important in this context because physical activity substantially declines from childhood through adolescence and is increasingly displaced by sedentary behavior, particularly screen time [6,7]. Data from the United Kingdom (UK), United States (US), and elsewhere show that less than 20% of adolescents meet the recommended 60 minutes per day of moderate-to-vigorous physical activity (MVPA) [8-11]. Active adolescents are more likely to become active adults and lead healthier lifestyles, gaining benefits both in the short-term (eg, bone health, mental health) and long-term (eg, sedentary behaviors, breast cancer, asthma, and self-esteem) [12,13]. A 2018 prospective cohort of 1826 UK adolescents (followed at 13, 14, and 15 years) found beneficial associations between device-based measures of MVPA and systemic metabolism (metabolic markers such as triglycerides, fatty acids, and systolic and diastolic blood pressure). Associations were more dependent on the most recent engagement, suggesting that regular physical activity sustains beneficial metabolic health and helps prevent disease [14].

Challenges in Changing Physical Activity in Adolescents

Despite the importance of promoting physical activity in adolescents, it is not yet clear what works. A recent meta-analysis of 17 cluster-randomized trials of school-based interventions, including children and adolescents up to 18 years old, found these did not affect accelerometer-measured MVPA [15]. Multidimensional interventions targeting physical activity across multiple settings, including school/home environments, policy, and parents, could promote, or at least prevent a decline, in physical activity in younger adolescents [15-17]. Despite the potential importance of parental support or involvement in multicomponent interventions [17], few studies have explored parental views, and the possibility of wide-scale implementation of such interventions is unclear. Digital interventions have been proposed as a solution to offer wide-reaching accessibility, but the majority of trials have been small, with mixed findings and web-based [18], which is not necessarily reflective of adolescent digital behavior.

Gaming Interventions to Change Physical Activity in Adolescents

Gaming constitutes a significant part of an adolescent's voluntary leisure time behavior, with some estimates of more than 90% playing for at least an hour per day [19]. Gaming is usually sedentary, but games that require bodily movement (ie, "exergames") have proved popular [20,21]. A 2015 meta-analysis of 35 trials showed that exergames, like Nintendo Wii, increased physical activity and improved physiological parameters as effectively as field-based physical activity and significantly enhanced enjoyment, self-efficacy, and intrinsic motivation for physical activity [22]. Although these small comparative trials were promising, no studies have explored exergaming as a population health intervention, which is likely to involve embedding exergaming in a larger multicomponent intervention that targets social and environmental determinants, including parental support [23-25].

Earlier generations of exergames such as Dance Dance Revolution released in 1998 and Nintendo's Wii Fit (2007), which sold over 22 million copies worldwide, and more recently, Pokémon Go, downloaded over 800 million times, are examples that have had huge commercial success [20,21]. Research studies on these earlier exergames have shown that playing exergames can increase energy expenditure up to 300% above resting levels and achieve physical activity of at least moderate intensity [22]. Randomized controlled trials in children [25,26], preadolescents [24], adolescents [27-29], and adults [30] found that exergames supported weight loss and increased fitness. However, trials have not been able to confirm exactly which psychological or social factors might lead to long-term engagement [31-33].

The Potential of Virtual Reality to Increase Physical Activity in Adolescents

Virtual reality has the potential to enhance the impact of exergaming by allowing the user to feel present and immersed in the virtual environment [34]. VR ownership is predicted to rise and estimated to be owned by most homes by 2022 [35]. With the possibility that VR will reach a wider population in the following years, VR hardware and software markets are reported to be increasing to 16 billion US dollars in 2022 from 6.2 billion in 2019 [36].

Small laboratory-based studies in adults have found that VR exergaming increases enjoyment and levels of physical activity but with lower perceived exertion than standard exercise conditions [37-39]. For example, one study involving 88 university staff and students found that although the heart rate was higher in VR than during regular exercise, participants reported feeling less tired and had higher ratings of enjoyment when VR was paired with exercise [40]. These results were echoed in an exploratory pre-post study in which 12 children engaged in VR-based biking (VirZoom) and traditional stationary exercise biking sessions for 20 minutes per session. The results showed no significant differences between the groups in the measures of heart rate. However, perceived exertion during the VR-based exercise was significantly lower, with participants also reporting significantly higher self-efficacy and enjoyment during the VR-based exercise compared to the traditional biking exercise session [38]. Another study of 30 people (between the ages of 6 and 50 years) playing a VR exergame for 15 minutes showed how perceived workout intensity correlated positively with the level of motivation and a significant increase in heart rate after gameplay [39]. These results were similar in a study of 60 females (18-30 years), which tested rowing performance, motivational, and affective impact during an aerobic exercise using VR and non-VR environments. VR groups rated physical activity tasks as more enjoyable, had improved performance (rowing longer distances, particularly the VR group which had a companion avatar) and did not perceive themselves to be exerting more physical effort when they did [37]. Thus, playing VR exergames with companions rather than alone raises and enhances physical activity levels even further. These psychological results, however, should be interpreted with caution due to small sample sizes and a failure to evaluate the effects of long-term physical activity.

Parental involvement is likely an important factor in the uptake of physical activity in younger adolescents [18]. We qualitatively interviewed 31 younger adolescents and determined that they were very interested in the concept of VR exergaming intervention, but that parental approval would be important for implementation [41]. To the best of our knowledge, no studies have explored the wider determinants, such as parental support, that would be required to implement a VR intervention.

The aim of this study, therefore, was to interview parents of adolescents (13-17 year-olds) to understand their views of physical activity, gaming, and head-mounted VR in order to gather evidence and build more understanding around this, in line with the Medical Research Council (MRC) Framework for developing complex interventions [42]. The MRC Framework includes several general stages, including planning, pilot and evaluation stages, reporting, and implementation [42].

Methods

This study formed part of the development work for the vEngage project [43]. Parents or primary caregivers with an adolescent child aged between 13 and 17 years were eligible to participate. They were recruited through social media, local secondary schools in London (UK), and the University College London (UCL) network and facilities (such as notice boards and online subject pool/study participation announcements) via emails and posters. The aim was to recruit up to 20 participants according to the recommended 10-20 interviews for a medium project using thematic analysis [44]. The study was approved by the UCL ethics committee (Project ID 12669/001), and all participants provided informed written consent before the interview.

Materials and Procedure

A semistructured interview schedule was developed to guide interviews (Multimedia Appendix 1). According to recommendations from Lewis and Ritchie [45], the schedule was designed to gather rich data on select topics including physical activity, gaming, VR, and the potential for using VR in a physical activity intervention. The schedule included open questions and specific probes to use if required. Interviews were conducted via telephone in June and July 2018 by one researcher (LM). The interviews were recorded and transcribed verbatim before analysis. As part of the interview schedule parents were asked whether their children had any disabilities that might impact their physical activity levels, but this data was not collected using a standardized questionnaire. Data was stored in accordance with the General Data Protection Regulation (GDPR).

Analysis

Framework analysis (a form of thematic analysis) was used to synthesize the data. Framework analysis is a systematic and effective approach for analyzing qualitative data in health research [46]. It is a recommended method for data that has been collected in semistructured interviews. Additionally, Framework analysis has been recommended for use when the data relates to a small number of topics and is appropriate for a sample of 10-20 participants [46].

Two researchers (LM and NF) independently analyzed three transcripts and generated a set of codes. The researchers then met to discuss, compare, and adjust the codes in order to develop the final analytical framework. The final framework was used to code all 18 transcripts by one researcher (LM). No particular software was used in coding the transcripts; however, the final set of themes and subthemes were collected with supportive quotes in an Excel spreadsheet (Microsoft). The data were then analyzed to explore emerging themes and identify common themes. Analyses were conducted with guidance from an experienced qualitative researcher (KN). The principal investigator (AF) independently reviewed the Framework Matrix and themes before the final interpretation.

Results

Participants were recruited through two local secondary schools in South London, UK (n=6), the University College London (UCL) network (n=8), and social media (n=4). Thirteen participants were mothers (age 48-58 years; mean age 53 years, SD 3 years), and five were fathers (age 48-58 years; mean age 52 years, SD 2 years) (see Table 1). Six had sons, two had daughters, and 10 had both. The mean age of all adolescents was 14 years (range 13-17 years, SD 1 years). Interviews lasted an average of 50 minutes (range 30-70 minutes). Only one adolescent had a disability (high-functioning autism spectrum disorder, Asperger syndrome), but this did not impact their ability to perform physical activity as reported by their parent. This difference was important to consider when analyzing our results because the literature on children and adolescents with autism spectrum disorders has shown difficulties in the development of coordination and movement [47]. However, the parent in question reported no issues in their child's coordination or movement. The final interpretation of the data is summarized in Table 2, with illustrative quotes in the text. Information in parentheses includes the child's age and gender.

Table 1. Demographic characteristics of parents and their adolescent children.

Characteristic	Participants (n=18), n	Parents' age (years), mean (SD)	Adolescents' age (years), mean (SD)
Mothers	13	53 (3.4)	14.8 (1.5)
Fathers	5	52 (2.1)	14.6 (1.4)

Table 2. Main themes and subthemes.

Main theme	Subthemes
General views on adolescent physical activity	• Strong belief in the importance of encouraging physical activity in adolescents
Views on gaming	• Adolescents game too often and this can be a dilemma • Addiction and violent content were major concerns • Accepting technology and gaming as facts of life and embracing potential • Peer influence and social pressure are strongest influences
General views on VR[a]	• Limited experience • VR gaming concerns • Concerns would be overcome if health benefits existed
VR exergaming	• Concern that a VR fitness game will be a fad • Regulations of use
Preference for real-world physical activity	• VR exergaming better than nothing • Obvious parental support may be off-putting • Harnessing screen time for a positive purpose strongly welcomed

[a]VR: virtual reality

General Views of Adolescent Physical Activity

Strong Belief in the Importance of Encouraging Physical Activity in Adolescents

Awareness of physical activity guidelines for adolescents was low. Only one participant knew that the adolescent physical activity guidelines were 60 minutes of MVPA per day. In total, 17/18 (94%) did not know the guidelines, and most were unaware that the guidelines existed. Some participants made accurate guesses.

Parents emphasized the importance of engaging adolescents in physical activity for lifetime health benefits. However, most specifically linked the importance of physical activity to their child's immediate mental health:

> *I think it is very [important] because I'm very, very active, and I know how good it is for your mind. (16F)*

> *I think anything that will get her moving more and doing more exercise will be good. (13F)*

Only two participants (11%) felt their child was engaged in enough physical activity. Most believed their child "should be doing more" (16F). All strongly believed that it was extremely important to encourage physical activity in adolescents:

> *Oh, it's massively important. I would put it one of the highest things, that and eating correctly. (13M)*

Views of Gaming

Adolescents Game too Often and This Can Be a Dilemma

All parents (100%) reported their child did some kind of gaming most days of the week, usually reflecting that they felt this was too often and this was associated with a level of guilt: "Probably six out of seven days a week" (16M) and "Every day and far too long. I feel guilty as a parent" (13M).

Addiction, Violent Content, and Time Spent Gaming Were Major Concerns

Nearly all participants (17/18, 94%) had negative opinions towards their child's gaming. All (100%) participants were concerned about gaming being violent. Most participants (14/18, 78%) reported trying to limit gaming, a sentiment generally raised by parents of younger adolescents:

> *Oh, I absolutely hate it. (14M)*

> *I do think they're incredibly addictive and then you get issues with trying to manage the time. (13M)*

> *I don't like the bloody, shooting kinds of things. I think some of it is a little too realistic. (16F)*

> *I wouldn't be letting him just have it in his bedroom overnight and that kind of thing. (M13)*

Participants did perceive some benefits to gaming, including skill development, such as becoming adept at using computers, developing visuospatial skills, and improving cognitive functioning. Another particularly common perceived benefit of gaming was social interaction:

> *A social aspect to it as well and even though it's not physical and in-person, there definitely is a lot of banter and a lot of chat, and they're very much in touch. I find that hugely beneficial. (13M)*

Accepting Technology and Gaming as Facts of Life and Embracing Potential

Participants felt that due to generational differences, they were unable to relate to gaming. However, 12/18 (67%) participants acknowledged that, when played in moderation, gaming could be fun:

> *I think it's just I find it hard to relate to because it's not something that I do. (13M)*

> *I think in moderation, it is a source of enjoyment. (13M)*

Unlike sedentary games, participants shared experiences of taking part in exergames with their children. Only 2/18 (11%) participants reported positive experiences with exergames such as the Wii Fit (Nintendo, 2006), but these were referred to as something done in the past. Despite concerns around gaming, participants accepted that gaming and technology were parts of life and could be used as a force for good. Participants felt that they could not fight against technology, that it was a reality of modern life:

> *It feels like an ancient device now, but at the time, it was one of those things that came along, and everybody was really excited about it. (13M)*

> *Technology is here, it's not going away, and that's fine if it has benefits for his health, then, yes, I'd always be up for that. (14M)*

Peer Influence and Social Pressure Are Perceived as the Strongest Influences in Adolescence

A majority of participants (10/18, 55%) highlighted the social pressure on adolescent gaming, particularly in terms of owning certain consoles and games. There was a general view that adolescents were influenced by peer and social pressure, then persuaded parents, rather than the reverse. Nearly all participants (17/18, 95%) underlined the role that the peer group plays in adolescent interests:

> *So that's the thing is to try and get something that becomes the really cool thing to do. (14M)*

General Views of Virtual Reality

Limited Experience

Most participants (15/18, 83%) had only a limited understanding of VR, and most had not tried it. A few had tried phone-based headsets. The same number of participants expressed an interest in VR, and some said that it seemed exciting. Others recognized it had potential:

> *It seems like you're actually in there, and you're moving within that space. And it's a whole lot more realistic, you're actually in the game rather than watching. (13M)*

> *I think it opens possibilities, definitely that's something I think that's really the future. (17F)*

Virtual Reality Gaming Concerns

A majority of participants (13/18, 72%) were more concerned about violent games in VR than in regular gaming:

> *Involving killing people, and that would not be appropriate in a virtual reality kind of scenario. (13M)*

> *It's quite a weird idea that you then have no idea what it is they're experiencing. (13M)*

Similarly, participants expressed safety concerns around multiplayer VR games as they may be unaware of whom their child is interacting with:

> *That she might be playing or getting in contact with people who she doesn't actually know. (13F)*

Further health and safety issues that participants mentioned concerned the practical matter of the space that VR would require.

> *I wouldn't want it taking over the sitting room. (16M)*

While participants suggested that gaming could be a social activity, they were concerned that VR gaming could be isolating:

> *It seems solitary, another solitary thing that, she'll be up in her room on her own. (13F)*

All parents cited cost as a barrier:

> *Obviously, there's the cost of it. (13M)*

Concerns Would Be Overcome if Health Benefits Existed

A third of the participants (6/18, 33%) suggested they would overcome their concerns if VR presented a tangible health benefit. Participants recognized that outside of gaming, VR could have the potential to be educational, and this was viewed as a positive thing:

> *If they (parents) can see the benefits of technology, then they're quite happy to invest in it. (13M)*

> *I'm sure there'll be some educational aspects to it that it could be used for as well. (13M)*

Virtual Reality Exergaming

Concern That a Virtual Reality Fitness Game Will Be a Fad

Participants did express concern that a VR fitness game could be a fad or novelty that comes and goes (7/18, 39%). However, participants suggested that if the game could maintain their child's interest, they would be more likely to invest:

> *It might be one of those fad things—they use it all the time for the first month, and then it'll die off slightly. (16F)*

> *Providing that it has some kind of stickiness in terms of it wasn't just a five-minute thing. (13M)*

Regulation of Use

A few participants (3/18, 17%) expressed that it would be important to regulate and manage the usage of the intervention. A majority of parents (13/18, 72%) expressed that they used technology and did not want to appear hypocritical:

> *It would need to be regulated really so that it's not just taking over. (16M)*

> *I think it's fine. I mean, we use technology for everything now, don't we? So it's an inevitable march. I've got no problem with it. I'd be a bit of a hypocrite if I did, considering I work in technology. (F14)*

Preference for Real-World Physical Activity

Virtual Reality Exergaming Better Than Nothing

Half of the participants (9/18, 50%) felt strongly that physical activity in the real world would provide greater benefits. Participants showed a preference for outdoor activities and team sports. Nearly all of the participants (17/18, 94%) thought their child would be excited to try a VR fitness game, and that they would embrace it:

> *I would see it as inferior to physical activity in the real world. (13M)*
>
> *I feel they're sitting indoors, on the computer, when they could be outside doing other things. (14M)*
>
> *They love new things, and they love the next step up technology-wise. He would love it, I'm sure. (13M)*

Obvious Parental Support May Be Off-Putting

However, 5/18 (28%) participants perceived that adolescents might be less keen to engage in something that their parents were actively encouraging. Participants also expressed the view there may be gender differences relating to a VR fitness game, such that boys may be more engaged than girls and that girls may engage differently.

> *I think that all teenagers and young people are developing a separation from their parents, and not necessarily wanting to do what their parents say. (13M)*
>
> *Maybe girls wouldn't be as into it as boys or maybe the time you have to give, you know, different challenges, different games. (14M)*

Participants also had ideas as to what may make the game appealing. These included levels, competition, a social element, and a challenge that requires skill:

> *I suppose some kind of score, so you could either beat people or beat your own score. (16F)*

Participants also reflected on the fact that adolescents expect the best in terms of quality and graphics:

> *They are tough consumers. If it's not on-brand, if it's not hitting their buttons, then they will just drop it and say, or won't even pick it up. (13M)*

Harnessing Screen Time for a Positive Purpose Strongly Welcomed

All participants welcomed the potential of gaming with a positive purpose, particularly mentioning being active during screen time:

> *If somebody has already lost them to screen time, having part of their screen time as exercise could be fantastic couldn't it? (13M)*

Discussion

Principal Findings

All participants in our study strongly believed that encouraging physical activity in adolescents was important. Most participants felt their children were not active enough. National datasets in adolescents using objectively measured physical activity suggest this perception is likely correct, with less than 15% meeting minimum guidelines [48]. Awareness of the recommended physical activity levels for adolescents was very low in our sample, mirroring the findings of our qualitative study in 31 adolescents [41] and a quantitative survey by our group in >1000 families showing that less than 20% of parents knew the recommended physical activity guidelines for their preschool children [49]. Parents who were aware of physical activity guidelines for their children were more likely to be supportive of physical activity [49], suggesting parental education ought to be incorporated in an intervention. Knowing targets for health would also assist with goal setting, a behavior change technique consistently associated with successful physical activity change [50].

Parents tended to have negative perceptions of gaming, particularly expressing concerns about addiction. Reflecting and perhaps exacerbating parents' concerns, "gaming disorder" is recognized in the International Classification of Diseases (ICD-11) [51], highlighting the scale and seriousness of addiction and possibly putting further pressure on parents to be alert to their children's gaming habits. The participants reported that trying to monitor their children's gaming and any digital intervention involving gaming would have to address this conflict. Positively using gaming time would be welcomed, so a possibility is that games designed in the future, used as physical activity intervention, could include materials and messages around replacing sedentary screen time with active gaming. A future game could also encourage breaks, and include real-world elements such as trying various sports or visiting clubs (eg, climbing wall, boxing, trampoline park). Parents in our study suggested that in an ideal world, their child would be encouraged to be active outdoors, but acknowledged that the gaming element would engage their child.

Parents reported they managed their child's gaming to some extent (usually with time limits), but acknowledged that adolescents should be allowed some autonomy, particularly as they approach young adulthood. This impression was supported by a study of 500 families using latent growth curve monitoring to demonstrate that parental media restriction decreases throughout adolescence [52]. Except for exergames, parents felt unable to relate to gaming, attributing this to a generational gap. The same effect was reported in a study of 80 teens aged 16-18 years and their parents exploring the mediation of mobile phone and internet use [53]. Exergaming was an exception in our study in terms of parental involvement. Almost half of the participants had used exergames, such as the Wii Fit in the past. Gamification of physical activity has been proposed as a way to encourage the integration of physical activity into their daily lives [54]. Those parents who tried exergames reported playing exergames with their children, and reported enjoying it (but recognizing that the technology was now outdated and was viewed as "past activity"). Playing with others was an important driver, including peer and social influence, and whether the game was believed to be "cool" [55].

Parents Had Limited Knowledge About Virtual Reality

Harnessing novel technology, such as the mobile game Pokemon GO, social media interventions and immersive VR games have been identified as an effective way of increasing physical activity in adolescents [54,56,57]. Therefore, the intervention should attempt to alleviate the discomfort that some experience when faced with new technology. For example, an information leaflet for parents enclosed in the intervention may help reduce parental concern.

Parents identified several benefits to gaming. These included socializing, becoming skilled at using technology, moving, or utilizing screen time for health benefits, the development of visuospatial skills and strategy. Participants were aware of VR but had a limited understanding of its use. Perceptions of VR were generally positive, with many participants describing it as interesting, intriguing, and as having potential. However, there was uncertainty around VR and its applications, and participants expressed concern about their child using violent games would be used in VR and have damaging psychological effects. Therefore, when developing the intervention, these factors should be considered, with violence avoided and, for example, providing guidance to help protect against addiction.

Parents reported they managed their child's gaming to some extent (usually with time limits) but acknowledged that adolescents should be allowed some autonomy, an effort supported by a study of 500 families using latent growth curve monitoring to demonstrate that parental media restriction decreases throughout adolescence [58]. Participants raised health and safety concerns, particularly around multiplayer games, with the belief that they may cause adolescents to interact with people they do not know. There was also concern about using a VR headset, causing injury or isolation for their child. Cost was a barrier reported by all participants. Many were not aware of the cost but believed it to be prohibitively expensive. The concern is not surprising given that VR is not yet mainstream technology, despite the predicted increase in household ownership by 2022 [36]. It is important that in the intervention development concerns are recognized and addressed by ensuring appropriate protective measures are taken, such as education around VR and safety.

Health Benefits of Using Head-Mounted Virtual Reality for Physical Activity Changed Parents' Attitude Toward Gaming

Despite concerns such as "novelty wearing off in 5 minutes," participants said they would support their children's use of VR if they knew there was a health benefit or educational element. Thus, despite reservations around VR, parents see the potential of the technology and would be open to it should it present a tangible health benefit. While some participants seemed to embrace technology and its potential for promoting health and fitness, others appeared merely accepting of technology and surrendering to its applications. These responses show that, while in different ways, all parents were open to using technology to improve the health of adolescents [58]. With the studies mentioned in the introduction, there is a potential risk of publication bias because studies showing positive or negative results are more likely to be published than those that show no results [18].

Parents reported a strong preference for real-world physical activity and felt that exercising using VR would be "better than nothing." There was a strong message for VR having the potential for adolescents who do little or no physical activity, which is the majority [59]. Therefore, an effective intervention is required, and a home-based physical activity intervention could particularly appeal to adolescents [6]. The intervention itself could also come with the direction that it is to be used in combination with outdoor exercise, which adolescents have previously suggested would be effective and appealing [41].

It may be important that the game is marketed directly to adolescents rather than to their parents since adolescents strive for independence [56]. Parents also suggested features they thought their children would find appealing in a game. Competition, levels, a challenge, and a social element were the aspects that came up most often and reflected findings from previous gaming research [60,61].

Only one adolescent child had a known disability, Asperger Syndrome. It was interesting to see that the mother of this child stated that she believed that gaming and head-mounted VR would help her child because "it takes away the pressure of having to interact with other people" (13M). Her feedback was interesting, but due to the aims of this study, we did not explore this further. While we recognize the potential for the benefit of exergames in children with disabilities, we felt that any disability that impacted movement might impact parent's perceptions of physical activity. Exergaming would best be excluded from the scope of our study.

Strengths and Limitations of the Study

The sample size was sufficient, according to Clarke and Braun's [41] recommendation of between 10 and 20 participants. It is not possible to say if the theoretical saturation was reached because we cannot be sure that the views of parents would differ in other populations [62,63]. Themes that reached saturation included all major themes, especially lack of experience with VR, general views on physical activity, worries around violent content, spending too much time online or gaming, and preferences for real-world physical activity.

Our sample comprised of more mothers than fathers and more parents of boys than girls, despite there being low levels of physical activity in both adolescent girls and boys. It would be useful to investigate whether parental concerns differ depending on the child's gender. Further research should also explore ways to engage boys and girls with the intervention equally. Social desirability bias may affect participants' answers since parenting methods can be a sensitive topic, but participants were open about how they knew their child should be engaged in more physical activity.

Conclusions

Adolescents have previously raised parental support as being an important factor in the intervention success [41], consistent with findings from a meta-analysis that found parental support correlated with adolescent physical activity [64].

The results of this study provide support for developing a head-mounted VR intervention to promote physical activity in adolescents. The findings provide useful insight for intervention development. Parents had negative associations with gaming but are accepting of it and embracing its potential. It is important that concerns are considered in the intervention development and negated, where possible, to maximize adoption and, ultimately, the efficacy of the intervention. Overall, parents believe encouraging physical activity in adolescents to be of importance. Therefore, while parents have reservations, it seems they would welcome anything that may improve their children's health, including if it involves harnessing a behavior like gaming and using it as a force for good. Recommendations for the next stage of intervention development would be to further research how best to educate and inform parents to reduce uncertainty around VR and the intervention. Additionally, shaping the game with an adolescent steering committee is recommended in order to ensure it is enjoyable and has longevity.

Acknowledgments

The vEngage study was funded by the Medical Research Council (Grant number MR/R015430/1). The authors are extremely grateful to all participating schools and young people and their families for their support. The authors would also like to thank their adolescent steering committee, who have helped shape the original idea for the study.

Conflicts of Interest

This research is sponsored by the Medical Research Council industry partnership grant in support of the development of a VR game licensed by Six to Start. There is no legal, financial, or commercial conflict with our industry partner company, Six to Start.

Multimedia Appendix 1
Interview schedule.
[DOCX File , 14 KB - games_v8i3e14920_app1.docx]

References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
3. Allender S, Foster C, Scarborough P, Rayner M. The burden of physical activity-related ill health in the UK. J Epidemiol Community Health 2007 Apr;61(4):344-348 [FREE Full text] [doi: 10.1136/jech.2006.050807] [Medline: 17372296]
4. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. The Lancet Child & Adolescent Health 2018 Mar;2(3):223-228. [doi: 10.1016/s2352-4642(18)30022-1]
5. Viner RM, Ozer EM, Denny S, Marmot M, Resnick M, Fatusi A, et al. Adolescence and the social determinants of health. Lancet 2012 Apr 28;379(9826):1641-1652. [doi: 10.1016/S0140-6736(12)60149-4] [Medline: 22538179]
6. Corder K, Sharp SJ, Atkin AJ, Griffin SJ, Jones AP, Ekelund U, et al. Change in objectively measured physical activity during the transition to adolescence. Br J Sports Med 2014 Apr 9 [FREE Full text] [doi: 10.1136/bjsports-2013-093190] [Medline: 24273308]
7. Hayes G, Dowd KP, MacDonncha C, Donnelly AE. Tracking of Physical Activity and Sedentary Behavior From Adolescence to Young Adulthood: A Systematic Literature Review. Journal of Adolescent Health 2019 Oct;65(4):446-454. [doi: 10.1016/j.jadohealth.2019.03.013]
8. Prevalence of insufficient physical activity. World Health Organization. URL: https://www.who.int/gho/ncd/risk_factors/physical_activity_text/en/ [accessed 2019-02-25] [WebCite Cache ID 76SHufrAr]
9. Thomson K. UK physical activity guidelines. Sport England. URL: https://www.nutrition.org.uk/attachments/504_5.UK%20Guidelines%20for%20physical%20activity%20-%20principles%20and%20subtleties_Kay%20Thomson.pdf [accessed 2019-02-25] [WebCite Cache ID 76SHzY9bq]
10. Craig R, Mindell J. Health Survey for England. Physical activity in children. URL: http://healthsurvey.hscic.gov.uk/media/63769/HSE2016-Child-health.pdf [accessed 2019-06-03] [WebCite Cache ID 78rbMjDEm]
11. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. The Lancet Child & Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
12. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273. [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]
13. Hallal PC, Victora CG, Azevedo MR, Wells JCK. Adolescent Physical Activity and Health. Sports Medicine 2006;36(12):1019-1030. [doi: 10.2165/00007256-200636120-00003]
14. Bell JA, Hamer M, Richmond RC, Timpson NJ, Carslake D, Davey Smith G. Associations of device-measured physical activity across adolescence with metabolic traits: Prospective cohort study. PLoS Med 2018 Dec;15(9):e1002649 [FREE Full text] [doi: 10.1371/journal.pmed.1002649] [Medline: 30204755]
15. Love R, Adams J, Sluijs EMF. Are school‐based physical activity interventions effective and equitable? A meta‐analysis of cluster randomized controlled trials with accelerometer‐assessed activity. Obesity Reviews 2019 Jan 09;20(6):859-870. [doi: 10.1111/obr.12823]

16. Sutherland R, Campbell E, Lubans DR, Morgan PJ, Okely AD, Nathan N, et al. 'Physical Activity 4 Everyone' school-based intervention to prevent decline in adolescent physical activity levels: 12 month (mid-intervention) report on a cluster randomised trial. Br J Sports Med 2015 Sep 10;50(8):488-495. [doi: 10.1136/bjsports-2014-094523]
17. Messing S, Rütten A, Abu-Omar K, Ungerer-Röhrich U, Goodwin L, Burlacu I, et al. How Can Physical Activity Be Promoted Among Children and Adolescents? A Systematic Review of Reviews Across Settings. Front. Public Health 2019 Mar 19;7. [doi: 10.3389/fpubh.2019.00055]
18. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A Systematic Review of Digital Interventions for Improving the Diet and Physical Activity Behaviors of Adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
19. Granic I, Lobel A, Engels RCME. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
20. Pocketgamer. biz. Pokemon Go captures 800 million downloads. URL: https://www.pocketgamer.biz/news/68209/pokemon-go-captures-800-million-downloads/ [accessed 2019-02-26] [WebCite Cache ID 76Tg6R0wJ]
21. Nintendo. co. Top Selling Title Sales Units. URL: https://www.nintendo.co.jp/ir/en/finance/software/3ds.html [accessed 2019-02-26] [WebCite Cache ID 76TeTDtAv]
22. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]
23. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). BMJ 2012 Sep 27;345:e5888. [doi: 10.1136/bmj.e5888] [Medline: 23044984]
24. Gao Y, Gerling K, Mandryk R, Stanley K. Decreasing sedentary behaviours in pre-adolescents using casual exergames at school. Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play. 2658693: ACM; 2014 Presented at: CHI-PLAY: Computer-Human Interaction in Play; October 2014; Toronto p. 97-106. [doi: 10.1145/2658537.2658693]
25. Lanningham-Foster L, Jensen TB, Foster RC, Redmond AB, Walker BA, Heinz D, et al. Energy expenditure of sedentary screen time compared with active screen time for children. Pediatrics 2006 Dec;118(6):e1831-e1835. [doi: 10.1542/peds.2006-1087] [Medline: 17142504]
26. Simon C, Kellou N, Dugas J, Platat C, Copin N, Schweitzer B, et al. A socio-ecological approach promoting physical activity and limiting sedentary behavior in adolescence showed weight benefits maintained 2.5 years after intervention cessation. Int J Obes (Lond) 2014 Jul;38(7):936-943 [FREE Full text] [doi: 10.1038/ijo.2014.23] [Medline: 24509504]
27. Staiano AE, Abraham AA, Calvert SL. Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention. Obesity (Silver Spring) 2013 Mar;21(3):598-601 [FREE Full text] [doi: 10.1002/oby.20282] [Medline: 23592669]
28. Staiano AE, Marker AM, Beyl RA, Hsia DS, Katzmarzyk PT, Newton RL. A randomized controlled trial of dance exergaming for exercise training in overweight and obese adolescent girls. Pediatr Obes 2017 Dec;12(2):120-128 [FREE Full text] [doi: 10.1111/ijpo.12117] [Medline: 26918815]
29. Dobbins M, De Corby K, Robeson P, Husson H, Tirilis D. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6-18. Cochrane Database Syst Rev 2009(1):CD007651. [doi: 10.1002/14651858.CD007651] [Medline: 19160341]
30. Huang H, Wong M, Lu J, Huang W, Teng C. Can using exergames improve physical fitness? A 12-week randomized controlled trial. Computers in Human Behavior 2017 May;70:310-316. [doi: 10.1016/j.chb.2016.12.086]
31. Krause JM, Benavidez EA. Potential Influences of Exergaming on Self-efficacy for Physical Activity and Sport. Journal of Physical Education, Recreation & Dance 2014 Mar 27;85(4):15-20. [doi: 10.1080/07303084.2014.884428]
32. Zeng N, Gao Z. Exergaming and obesity in youth: current perspectives. Int J Gen Med 2016;9:275-284 [FREE Full text] [doi: 10.2147/IJGM.S99025] [Medline: 27536158]
33. Sun H. Impact of exergames on physical activity and motivation in elementary school students: A follow-up study. Journal of Sport and Health Science 2013 Sep;2(3):138-145. [doi: 10.1016/j.jshs.2013.02.003]
34. Slater M, Wilbur S. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. Presence: Teleoperators and Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
35. Statista. com. Virtual Reality (VR) - Statistics & Facts. URL: https://www.statista.com/topics/2532/virtual-reality-vr/ [accessed 2019-02-26] [WebCite Cache ID 76TjvDfKu]
36. Liu S. Statista. Virtual Reality (VR) - statistics & facts.: Statista; 202- Jun 23. URL: https://www.statista.com/topics/2532/virtual-reality-vr/ [accessed 2020-08-12]
37. Murray EG, Neumann DL, Moffitt RL, Thomas PR. The effects of the presence of others during a rowing exercise in a virtual reality environment. Psychology of Sport and Exercise 2016 Jan;22:328-336. [doi: 10.1016/j.psychsport.2015.09.007] [Medline: 30872122]

38. Zeng N, Pope Z, Gao Z. Acute Effect of Virtual Reality Exercise Bike Games on College Students' Physiological and Psychological Outcomes. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):453-457. [doi: 10.1089/cyber.2017.0042] [Medline: 28715263]
39. Finkelstein S, Suma EA. Astrojumper: Motivating Exercise with an Immersive Virtual Reality Exergame. Presence: Teleoperators and Virtual Environments 2011 Feb;20(1):78-92. [doi: 10.1162/pres_a_00036]
40. Plante TG, Aldridge A, Bogden R, Hanelin C. Might virtual reality promote the mood benefits of exercise? Computers in Human Behavior 2003 Jul;19(4):495-509. [doi: 10.1016/S0747-5632(02)00074-2]
41. Farič N, Yorke E, Varnes L, Newby K, Potts HW, Smith L, et al. Younger Adolescents' Perceptions of Physical Activity, Exergaming, and Virtual Reality: Qualitative Intervention Development Study. JMIR Serious Games 2019 Jun 17;7(2):e11960 [FREE Full text] [doi: 10.2196/11960] [Medline: 31210135]
42. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
43. 2020. vEngage Study. 2019 Jan 20. URL: https://vengagestudy.com [accessed 2020-08-12]
44. Clarke V, Braun V. Successful qualitative research: A practical guide for beginners. Los Angeles: Sage; 2013.
45. Lewis J, Ritchie J. Qualitative research practice: A guide for social science students and Researchers. Los Angeles: Sage; 2003.
46. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
47. Srinivasan S, Pescatello L, Bhat A. Current perspectives on physical activity and exercise recommendations for children and adolescents with autism spectrum disorders. Phys Ther 2014 Jun;94(6):875-889 [FREE Full text] [doi: 10.2522/ptj.20130157] [Medline: 24525861]
48. Craig R, Mindell J. Health Survey for England. Physical activity in children. 2017 Dec 13. URL: http://healthsurvey.hscic.gov.uk/media/63769/HSE2016-Child-health.pdf [accessed 2020-08-12]
49. Sawyer A, Smith L, Schrempft S, van Jaarsveld CH, Wardle J, Fisher A. Primary caregiver knowledge of paediatric physical activity recommendations in the United Kingdom and its association with caregiver behaviour: an observational study. BMC Public Health 2014 Aug 04;14:795 [FREE Full text] [doi: 10.1186/1471-2458-14-795] [Medline: 25092511]
50. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
51. Available at:. International Classification of Diseased (ICD-11). URL: https://icd.who.int/browse11/l-m/en [accessed 2019-06-03] [WebCite Cache ID 78rfP4YVd]
52. Padilla-Walker LM, Coyne SM, Fraser AM, Dyer WJ, Yorgason JB. Parents and adolescents growing up in the digital age: latent growth curve analysis of proactive media monitoring. J Adolesc 2012 Oct;35(5):1153-1165. [doi: 10.1016/j.adolescence.2012.03.005] [Medline: 22507383]
53. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
54. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. Prev Med Rep 2015;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
55. Trepte S, Reinecke L, Juechems K. The social side of gaming: How playing online computer games creates online and offline social support. Computers in Human Behavior 2012 May;28(3):832-839. [doi: 10.1016/j.chb.2011.12.003]
56. Fisher A, Meisel S, Webb T, Fisher B. Innerselfie: can we use Immersive Virtual Reality to change cancer risk perceptions and promote physical activity in young people? Conference Abstract: 2nd Behaviour Change Conference: Digital Health and Wellbeing 2016. [doi: 10.3389/conf.fpubh.2016.01.00086]
57. LeBlanc AG, Chaput J. Pokémon Go: A game changer for the physical inactivity crisis? Prev Med 2017 Aug;101:235-237. [doi: 10.1016/j.ypmed.2016.11.012] [Medline: 27856340]
58. Sanders W, Parent J, Forehand R, Sullivan ADW, Jones DJ. Parental perceptions of technology and technology-focused parenting: Associations with youth screen time. J Appl Dev Psychol 2016;44:28-38 [FREE Full text] [doi: 10.1016/j.appdev.2016.02.005] [Medline: 27795603]
59. 2009. NHS Digital-Health Survey for England 2008: Physical activity and fitness. URL: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2008-physical-activity-and-fitness [accessed 2019-02-25] [WebCite Cache ID 76SI9jwOY]
60. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
61. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]

62. Fugard AJ, Potts HW. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. International Journal of Social Research Methodology 2015 Feb 10;18(6):669-684. [doi: 10.1080/13645579.2015.1005453]
63. Curtis B, Curtis C. Social research: A practical introduction. Los Angeles: Sage; 2011.
64. Yao CA, Rhodes RE. Parental correlates in child and adolescent physical activity: a meta-analysis. Int J Behav Nutr Phys Act 2015 Feb 11;12:10 [FREE Full text] [doi: 10.1186/s12966-015-0163-y] [Medline: 25890040]

Abbreviations

ASD: autism spectrum disorder
GDPR: General Data Protection Regulation
MRC: Medical Research Council
MVPA: moderate-to-vigorous physical activity
VR: virtual reality

Edited by G Eysenbach; submitted 03.06.19; peer-reviewed by K Ng, J Salisbury; comments to author 29.09.19; revised version received 19.11.19; accepted 25.06.20; published 25.08.20.

Please cite as:
McMichael L, Farič N, Newby K, Potts HWW, Hon A, Smith L, Steptoe A, Fisher A
Parents of Adolescents Perspectives of Physical Activity, Gaming and Virtual Reality: Qualitative Study
JMIR Serious Games 2020;8(3):e14920
URL: http://games.jmir.org/2020/3/e14920/
doi:10.2196/14920
PMID:32840487

Viewpoint

Quality Criteria for Serious Games: Serious Part, Game Part, and Balance

Polona Caserman[1], MSc; Katrin Hoffmann[2], Dip (Sportwiss); Philipp Müller[1], MSc; Marcel Schaub[3]; Katharina Straßburg[1]; Josef Wiemeyer[2], Prof Dr rer med; Regina Bruder[3], Prof Dr; Stefan Göbel[1], PD Dr Ing

[1]Multimedia Communications Lab, Technical University of Darmstadt, Darmstadt, Germany

[2]Institute of Sport Science, Technical University of Darmstadt, Darmstadt, Germany

[3]Research Group Didactics of Mathematics, Technical University of Darmstadt, Darmstadt, Germany

Corresponding Author:
Polona Caserman, MSc
Multimedia Communications Lab
Technical University of Darmstadt
Rundeturmstraße 10
Darmstadt, 64289
Germany
Phone: 49 6151 16 20391
Email: polona.caserman@kom.tu-darmstadt.de

Abstract

Serious games are digital games that have an additional goal beyond entertainment. Recently, many studies have explored different quality criteria for serious games, including effectiveness and attractiveness. Unfortunately, the double mission of serious games, that is, simultaneous achievement of intended effects (serious part) and entertainment (game part), is not adequately considered in these studies. This paper aims to identify essential quality criteria for serious games. The fundamental goal of our research is to identify important factors of serious games and to adapt the existing principles and requirements from game-related literature to effective and attractive serious games. In addition to the review of the relevant literature, we also include workshop results. Furthermore, we analyzed and summarized 22 state-of-the-art serious games for education and health. The selected best-practice serious games either prove their effectiveness through scientific studies or by winning game awards. For the analysis of these games, we refer to "DIN SPEC 91380 Serious Games Metadata Format." A summarized text states quality criteria for both the serious and the game part, and especially the balance between them. We provide guidelines for high-quality serious games drawn from literature analysis and in close cooperation with domain experts.

(JMIR Serious Games 2020;8(3):e19037) doi:10.2196/19037

KEYWORDS

serious games; educational games; games for health; exergames; quality criteria; video games

Introduction

Serious games are digital games that not only entertain but also intend to achieve at least one additional goal, a so-called characterizing goal [1]. Serious games aim to accomplish this characterizing goal without compromising the experience of playing a game (player experience). Examples include educational games, such as Meister Cody – Talasia (Meister Cody GmbH) [2], and games for health, such as ExerCube (Sphery AG) [3] and Pokémon GO (Nintendo Co) [4]. They should, in general, motivate the player to achieve the characterizing goal through appropriate methods, an engaging player experience, and the use of appropriate interaction technology.

Serious games are not limited to training (exercising) or learning. They can also be used for other purposes and can be applied in almost every area. For example, America's Army (US Army) [5] is used as a recruiting tool; Re-Mission (HopeLab) [6,7] aims to change attitude, knowledge, and behavior; Trash Monsters (Bunny and Gnome) [8] improves knowledge about waste separation; and Orwell: Keeping an Eye on You (Osmotic Studios) [9] raises awareness about government surveillance. In some cases, games that have not been intentionally developed for serious purposes can also achieve additional effects. For example, the Civilization (Take-Two Interactive Software) [10], Age of Empires (Microsoft Corp) [11], and Assassin's Creed (Ubisoft) [12] series are primarily developed for entertainment purposes;

however, when playing these games, players also learn about ancient organizations or the history of civilizations. Nevertheless, the question remains: what are important constituents of serious games and, in particular, what are important aspects of high-quality serious games? Although many useful heuristics are presented in the game-related literature [13,14], no model yet exists that focuses equally on the serious and game aspects, as well as on the balance between them.

Existing quality criteria for video games often focus on appropriate game design, (eg, they consider only the player enjoyment [13] and are identified based on game reviews and rating systems [14,15]). Rating systems cover the different kinds of expertise of involved game reviewers and the complexity of testing processes [16]. For example, IGN (IGN Entertainment), Gamespot (CBS Interactive), and PC Gamer (Future US Inc) primarily rely on the expertise and opinions of their in-house editors. The metareview system Metacritic (CBS Interactive) aggregates these expert scores into a single metascore, in addition to letting users vote on a secondary user score. Conversely, studies that evaluate quality criteria for serious games are often specific to an application field and focus on didactic aspects (ie, they propose specific guidelines for educational [17] or motion-based serious games [18]). In particular, existing quality criteria for serious games often lack in the combining of both aspects (ie, serious as well as game aspects).

In this paper, we aimed to gather criteria for high-quality serious games, considering both the serious and game aspects and the balance between them. First, we described some of the successful serious games according to the proven “DIN SPEC 91380 Serious Games Metadata Format” (SG-MDF) [16,19]. Existing serious games taxonomies include specific classification systems for rehabilitation games [20], educational games [21,22], or serious games in general [23-25]; however, they usually select classification criteria arbitrarily and are not generally accepted. In particular, they do not take the aspects of the characterizing goal or the quality of serious games into consideration. SG-MDF overcomes these limitations and covers all crucial aspects of serious games (ie, the characterizing goal as well as quality criteria). Finally, based on the description of serious games, relevant game-related work, and close cooperation with domain experts, we refined and extended these quality criteria to define high-quality serious games.

To summarize, our primary goal is to identify quality criteria for the serious and game part, as well as the balance between them. We have provided guidelines for high-quality serious games drawn from literature analysis and workshops with domain experts.

State of the Art

Game Classification and Selection

We studied serious games to classify them according to SG-MDF. SG-MDF is also used in a metadata-based information system [26], which allows providers of serious games (eg, developers and publishers) to systematically describe the games so that users (eg, teachers, trainers, coaches, doctors, and therapists) can find suitable games accordingly. Using this format, we provide a summary of games for health [3,4,6,7,27-52] and educational games [2,8,9,53-64] (see Multimedia Appendices 1 and 2). Note that the list is not complete and should only serve as a foundation for developing and explaining the quality criteria. We selected serious games that prove their effectiveness either through scientific studies or by winning game awards. Furthermore, the selected serious games have a certain level of familiarity in the community.

We use SG-MDF because it overcomes the limitations of the existing taxonomies [23-25] and covers crucial aspects of serious games, such as the characterizing goal and the quality (based on scientific studies, game awards, professional ratings, recommendation by experts, and number of players/downloads). However, we did not include all categorizations as proposed by SG-MDF and included only measures that are important to present the quality criteria in this paper. For example, game modes are important for the “support social interactions” criterion, and the target group is essential for the “appropriate interaction technology” and “media presentation” criteria. In general, all serious games should use appropriate interaction technology for the target group, game purpose, and application area. Furthermore, the progress indicator is essential for the “appropriate feedback and reward” criterion.

Games for Health

Serious games are not only fun to play but are also beneficial for health. For example, they can motivate players to increase physical exercise. Due to insufficient physical activity, the risk of diseases such as obesity, diabetes, cancer, and cardiovascular diseases are increasing. The World Health Organization reports that physical activity has decreased over time in high-income countries [65]. These results show that it is crucial to motivate people to become more physically active. However, games for health do not only cover physical exercises but are also often used for prevention, rehabilitation and, in general, supporting health (ie, enforcing a behavior change towards a better, more active, and healthier lifestyle, including better nutrition).

Popular exergames such as Pokémon GO [27] and Dance Dance Revolution (Konami) [28] aim to provide an effective and attractive workout experience for a wide variety of users. Pokémon GO, for example, has over 1 billion downloads on Google Play Store (Google Corp) [44], making over $800 million US dollars worldwide in 2019 [45]. The study by Althoff et al [4] shows that it indeed increases players’ activity level compared with their prior activity level; however, the researchers could only confirm short-term effects. Additional studies show that Dance Dance Revolution significantly increases energy expenditure [29] and improves aerobic fitness in overweight children [30]. Some schools have even included the game in their physical education courses to motivate children to exercise [66]. However, studies also report that exergames are often only capable of providing light to moderate exercise and thus, often fail to significantly increase physical activity or exercise attendance [67].

Furthermore, Wii Sports games (Nintendo Co) are best-selling video games [68] that contribute to weight loss [31] or to increased muscle strength [32]. However, they lack proper training concepts and disregard performance aspects that are essential for a successful workout (eg, accuracy, precision, and intensity of movement). Similarly, Beat Saber (Beat Games) [33] is one of the top virtual reality games in 2019 [46,47]; however, it focuses on game design and ignores the extensive knowledge of movement and training science in sports. On the contrary, the ExerCube [34] was developed by an interdisciplinary team of sports scientists, game designers, and researchers in the field of human-computer interaction. The results of a user study with 40 participants show that the ExerCube is on par with personal training [3].

Other serious games intend to improve the physical status of older people. For example, BalanceFit aims to improve coordination, strength, and balance [35]. A study by Hardy et al [36] shows that an adaptive approach enables people with heterogeneous skills to play this game (eg, fit players as well as players with gait impairments or wheelchairs). Furthermore, the game ErgoActive provides adaptive cardio training on an ergometer bike to increase the physical activity of its players [35]. The results of a feasibility study with 16 participants demonstrated the effectiveness of cardio training based on personalized heart rate control [37].

The characterizing goal of a game for health does not necessarily need to aim at a physical training effect. The serious game Re-Mission [38] intends to inform patients about cancer treatments and aims to change their attitude positively. Studies confirmed the effectiveness of the game in randomized controlled trials with cancer patients [6,7]. Other serious games, such as Escape from Diab and Nanoswarm: Invasion from Inner Space (Archimage) are persuasive and are able to change health-related behavior among children [39]. Further serious games for health are used as prevention (eg, PlayForward: Elm City Story) [40] or rehabilitation (eg, SnowWorld) [41]. Moreover, Dr Kawashima's Brain Training (Nintendo Co) includes a set of minigames that are designed to improve cognitive functions in elderly persons. However, even though randomized controlled trials report benefits [42,43], no long-term effects and relevance for everyday functioning could be confirmed.

Educational Games

In addition to motivating players to become more physically active, serious games are often used to increase players' motivation levels to learn and improve learning outcomes. Educational games can be a reliable and effective tool compared with traditional methods [69]. According to the Entertainment Software Association, 74% of parents believe that video games can be educational for their children [70]. Educational games are also the second most popular Google Play app category [71]. Serious games are effective in terms of learning and some of them are even better at teaching than traditional methods [72].

Studies in the field of game-based learning show the benefits of educational games, including improvement of mathematical skills (eg, Meister Cody – Talasia [2]), reading performance (eg, Meister Cody – Namagi [53]), and programming skills (eg, Debugger 3.16: Hack'n'run [Spiderworks Games] [54,61]). Serious games can also be used to assess knowledge (eg, Semideus [Flow Factory] [55,56]). VocabiCar (Westermann Digital GmbH) [57] is another educational game for children and intends to expand the English vocabulary of pupils.

In addition to improving players' skills, educational games can also raise awareness. The game Trash Monsters [8] raises awareness of waste separation and teaches children how to recycle correctly. Educational games do not necessarily need to be intended for children but can be dedicated to students or adults in general. Orwell: Keeping an Eye on You serves to raise awareness of state surveillance [9,63], and Orwell: Ignorance is Strength (Osmotic Studios) serves to raise awareness of fake news [58,64]. By representing different moral values, the game strengthens or enforces players' opinions. However, both serious games are reading intensive and therefore (similar to all serious games) only suitable for specific player types.

Another application area for serious games are simulations, in particular, corporate games for training purposes. ViPOL (TriCAT) enables police forces to train in virtual reality for scenarios that are too expensive, complex, or dangerous to be trained for in the real world [59]. The simulation was developed in close cooperation with police officers. In a study by Bertram et al [59], the results show that virtual training can be as efficient as regular training for complex collaborative tasks.

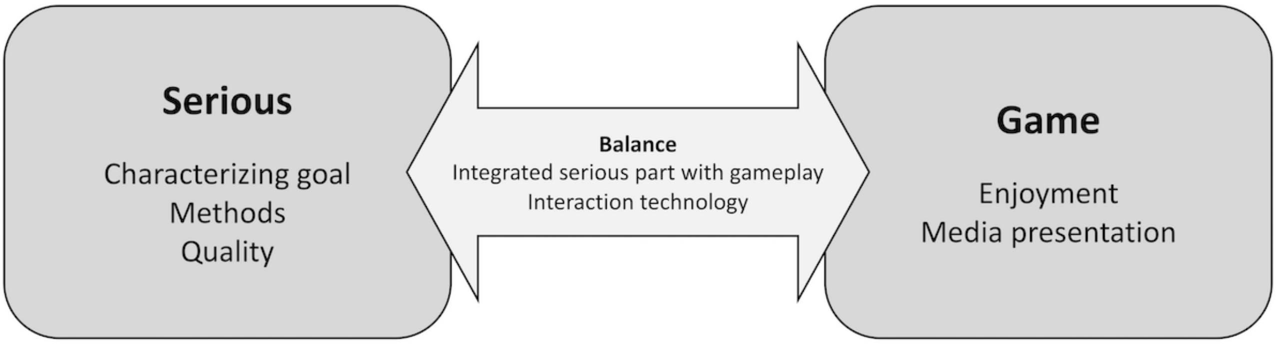
Identifying Quality Criteria for Serious Games

Development of the Criteria

A review of the state of the art was conducted to determine quality criteria for attractive and effective serious games. We propose essential aspects of high-quality serious games, including characteristics for the serious and game part, as well as for the balance between them. Although we focus on educational games and games for health, the quality criteria are transferable to all kinds of application areas. For example, the criteria can be used not only for games that improve players' skills/performance but also for games that raise the players' awareness of a certain topic or that positively change their attitudes.

We furthermore discussed the derived quality criteria in workshops with experts from the respective areas. The aim of the workshops was to identify the requirements and needs for high-quality serious games. Therefore, domain experts, such as game developers and companies that deploy serious games, as well as scientists from different areas (eg, sports education and computer science), critically discussed the quality criteria of serious games. The quality criteria are shown in Figure 1 and further detailed in the following section. High-quality serious games must achieve both the serious and the game aspects; they must systematically support players to reach the characterizing goal (serious part) and they must elicit and maintain player experience (game part). Furthermore, both parts should be perfectly matched and integrated rather than addressed in isolation.

Figure 1. Quality criteria for the serious part and game part, as well as the balance between them.



In the following section, we summarize key findings, including the strength of evidence for each criterion. These results should help other researchers and game developers gain a deeper understanding of high-quality serious games.

Serious Part

This section describes the core elements for the serious part of the game: existence of a characterizing goal, development of appropriate methods for achieving this characterizing goal, and evaluation of the quality (see also Table 1).

Table 1. Summary of quality criteria for the serious part.

Quality criteria and relevant quality aspects	Explanation
Characterizing goal	
Focus on the characterizing goal	• Learning/training goal must remain in focus, for which a combination of physical and cognitive training can be beneficial • Support players to achieve the characterizing goal • Game elements should not interfere with the learning/training process
Clear goals	• Appropriate methods for the specific application area and target group • Goals are clear and appropriate so that players can work towards the characterizing goal
Indispensability of the characterizing goal	• Serious part must be mandatory • Characterizing goal must not be avoidable • Training and learning tasks should not be a hurdle
Methods	
Correctness of the domain expert content	• Avoid errors and ensure that the content is technically correct • Ensure correct technical language • Remain neutral, especially on political and social issues
Appropriate feedback on progress	• Players should receive feedback on their performance and progress • Visible and recognizable effects • Provide simultaneous feedback (eg, visual, audio, haptic; multimodal feedback)
Appropriate rewards	• Provide positive reinforcement and in-game awards
Quality	
Proof of effectiveness & sustainable effects	• Prove that the characterizing goal is achieved • Learning/training effects need to be sustainable
Awards and ratings	• Game awards, professional and user ratings, recommendations by domain experts, game reviews, and number of players/downloads state the quality of the game

Characterizing Goal

Serious games must ensure that players achieve the characterizing goal.

Focus on the Characterizing Goal

The characterizing goal of a serious game is closely linked to the application area. In educational games, the characterizing goals include learning or training effects. In games for health, the characterizing goals include changes in vital status or general changes in attitude and behavior (eg, nutrition or mobility behavior). It may also be beneficial to combine physical and cognitive training. For example, in the game Lü (Lü Interactive Playground) [73], the players not only improve mathematical skills but also stay physically active. Similarly, ExerCube [34] provides training for body and mind. Furthermore, recent evidence on exergame-based therapy for Parkinson disease shows that exergames can enhance cognitive skills and are at least as effective as traditional therapies [74,75].

Thus, serious games should always focus on achieving the characterizing goal and should support the player in achieving this goal. Learning or training content must remain in focus during gameplay and game elements should not interfere with the learning or training process.

Clear Goals

Similar to entertainment games, in which the game goals should be clear [13,14], the characterizing goal of a serious game should also be transparent so that players can work towards achieving this goal. In particular, a serious game should ensure that players always know what to do to complete the tasks or exercises; otherwise, a tutorial is required (see also the "Intuitive Game Mechanics and Natural Mapping" section). For example, in the ExerCube, a virtual avatar demonstrates required movements in a short tutorial to make sure that players know how to execute them [3]. The educational games Meister Cody – Talasia and Meister Cody – Namagi [60] show how each exercise should be solved before players start to solve the tasks.

Indispensability of the Characterizing Goal

Engaging in the serious part of the game should be mandatory for playing the game. Otherwise, players may avoid the serious part to get to the fun part more quickly [76], preventing the characterizing goal from being reached. In particular, the training and learning tasks should not be a hurdle while playing a serious game. In other words, the characterizing goal should be embedded in the gameplay (see also the "Integrated Serious Part With Gameplay" section).

Methods

The methods for serious games need to be appropriate for the specific application area and target group.

Correctness of the Domain Expert Content

The most evident requirement for serious games is that they must not contain any errors with respect to their subject matter, such as erroneous mathematical equations, incorrect information on historical events, or inadequate information on physical exercise. If typing or presenting errors occur, these errors must not mislead players. Furthermore, conveyed information must not only be factually correct but also be imparted using appropriate technical language.

Moreover, even though there is usually no connection between neutrality and the correctness of a given opinion or issue, serious games should remain neutral, especially on political and social issues. For example, the games Orwell: Keeping an Eye on You [9] and Orwell: Ignorance is Strength [58] do not convey a specific political opinion to the player. Instead, they only show the effects of the player's actions without judging them. The player then has the opportunity to assess and question their own decisions. In particular, serious games should be appropriate for the target group, depending on religion, culture, and traditions.

Appropriate Feedback on Progress

A serious game should provide appropriate feedback to players so that they can assess their progress. Thus, to enhance player performance, effects should be visible and recognizable (eg, through a progress bar). Continuous feedback on progress is essential in all serious games as the players work towards achieving the characterizing goal [77]. Moreover, multimodal feedback (eg, visual, audio, or haptic feedback) can be beneficial [14]. For example, in the ExerCube, players receive immediate visual and audio feedback to enhance their movements [3].

Apart from in-game feedback, Ravyse et al [78] furthermore show that postgame feedback also improves learning. Similarly, the level number can indicate the player's progress (ie, the higher the level, the more skills the player has developed). A higher in-game level seems to have a more significant effect on motivation (desire to practice) than individual rewards (eg, achievements) [79]. Game statistics additionally show the player's progress after ending the game or a level and are not only advantageous for players themselves, but also useful for their therapists (especially in games for health [20]) or their parents and educators (in educational games [60]).

Appropriate Reward

Games should provide positive reinforcement and in-game awards [1] to immerse players more deeply in the game [14]. For example, in the educational game VocabiCar, players who accomplish a learning task or complete a challenge gain points and can access their progress in a high-score table [57]. High-score tables allow players to compare their performance against other players. In addition to points, in-game awards consist of virtual badges, achievements, power-ups, and desirable objects. Fancy animations or a possibility to change their avatar (eg, new clothes, hair color, or equipment) as a reward for accomplishing a task can further motivate players. However, there is ample evidence that "rewards or feedback delivered in a controlling manner undermine intrinsic motivation and deeper forms of learning" [80]. Therefore, rewards should be deliberately deployed in serious games.

Quality

High-quality serious games should measure the effects and benefits in a scientific study. Furthermore, awards and user or domain expert ratings can confirm the quality of a serious game.

Proof of Effectiveness and Sustainable Effects

A serious game is effective when players achieve the characterizing goal and the learning or training effects are sustainable. Researchers often validate effectiveness with a study, such as a scientific, clinical, or empirical evaluation, by monitoring heart rate (eg, ExerCube [3] and ErgoActive [37]), number of steps (eg, Pokémon GO [4]), or aerobic fitness (eg, Dance Dance Revolution [30]). However, potential aversions to video games, certain game genres, or specific interaction technologies among players have to be considered when designing and evaluating a study. In educational games, the results of a group exposed to a serious game and a group exposed to traditional methods can be compared (eg, Meister Cody – Talasia [2] and ViPOL [59]). However, empirical studies often suffer from numerous sources of error (eg, bias) [81]. In this regard, randomized controlled trials with a sufficient number of participants are the gold standard for empirical proof of effects.

Awards and Ratings

In addition to scientific studies that evaluate usability and player experience, game awards (eg, German Computer Games Award, European Innovative Game Award, and International Educational Games Competition) are also an important aspect of identifying high-quality serious games. Further quality criteria include professional or user ratings, the number of players or downloads (eg, Google Play, App Store [Apple Inc], and Steam [Valve Corp]), and recommendations by domain experts and game reviews (eg, IGN, GameSpot, and PC Gamer).

Game Part

This section describes core elements for appropriate game design and suitable interaction technology, as seen in Table 2. Note that there is considerable overlap in the various concepts within the game part.

Table 2. Summary of quality criteria for the game part.

Quality criteria and relevant quality aspects	Explanation
Enjoyment	
Ensure player engagement and experience	• Ensure positive experience during playing • Serious games should be engaging and enjoyable (Koster's theory of fun for game design [82], GameFlow approach [13], and PLAY[a] heuristics [14]) • Provide an engaging experience for different player types
Ensure flow	• Keep a balance between a player's skills and challenge (Csikszentmihalyi's flow theory [83]) • Dynamically adapt the difficulty level depending on the current player's performance in the game • Adapt to players to increase effectiveness (eg, motivate them to repeat the exercises continuously and regularly) • Increase complexity as the player gets better (Bushnell's theorem of "easy to learn, difficult to master" [84]) • Provide varied gameplay
Establish an emotional connection	• Allow emotions and arouse instinct (Dillon's 6-11 framework [85], LeBlanc's theory of 8 kinds of fun [86])
Sense of control	• Players should have control over their actions in the game
Support social interactions	• Provide different game modes (collaborative and competitive settings for players that perform better in groups)
Ensure immersive experience	• Include multimodal sensory stimulations: visual, audio, haptics, smell • Ensure the sense of "being there"
Media presentation	
Attractive graphics	• Graphics must be appropriate for the game purpose, application area, and target group • Ensure clear interface without unnecessary information to not distract players from a specific task
Appropriate sound	• Include appropriate background music and sound effects

[a]PLAY: Heuristics of Playability.

Enjoyment

Serious games should not only ensure positive player experience, flow, and sense of control but should also support social interaction.

Ensure Player Engagement and Experience

Player engagement is tightly associated with enjoyment. Koster, the author of *A Theory of Fun for Game Design*, addresses the importance of a game being engaging, enjoyable, and fun [82]. The GameFlow approach proposed by Sweetser and Wyeth [13] includes 8 dimensions with numerous criteria and recommendations to ensure player enjoyment in games. Calvillo-Gámez et al [87] furthermore present the core elements of the gaming experience to provide a positive experience while playing video games. Moreover, Desurvire and Wiberg [14] propose heuristics of playability for game developers to develop better games.

However, due to different kinds of players, not every player will find all components equally important. Therefore, the game should provide different fun components to provide an engaging experience for different player types (eg, Bartle's player types [88]).

Ensure Flow

For the optimal player experience, the game has to establish a satisfying balance between challenges and skills.

Csikszentmihalyi's well-known flow theory describes the feeling of enjoyment when the task difficulty and skill levels are in balance [83]. This theory is complemented by Bushnell's theorem of "easy to learn, difficult to master" [84]. Games that are easy to learn enable flow because they are not overwhelming, whereas games that are hard to master keep players from dropping out because of boredom. Thus, as the player's performance improves, the complexity or difficulty of the game should also increase. In other words, serious games should adapt to the current performance level of the player (ability vs skills).

To avoid boredom, game developers should also ensure that the gameplay varies. As proposed by Desurvire and Wiberg [14], any fatigue or boredom should be minimized by varying activities and pacing during the gameplay. Furthermore, research by Scoresby and Shelton [89] identified that content, emotion, motivation, and engagement associated with the game are necessary criteria for flow.

Moreover, serious games should automatically adapt to the players to motivate them to keep learning/training and to increase the effectiveness. One of the primary advantages of educational games is their ability to engage the learner so that they voluntarily complete sufficient repetitions of activities, ensuring that learning takes place [90]. For example, both of the Meister Cody educational games adapt their difficulty depending on the player's skills [60]. The results show that, due to adaptivity, game-based learning is particularly promising for children who want to learn in a home environment or do not have access to individual reading support [53]. Similarly, exergames should ensure that the intensity matches the player's fitness level in order to motivate players to repeat the exercises continuously and regularly [91]. For example, ExerCube [3] or ErgoActive [37] identify the individual's optimal strain to adapt the game difficulty and complexity gradually based on the player's heart rate.

Establish an Emotional Connection

Additionally, players should get emotionally involved in a serious game. Dillon [85] has drawn attention to the fact that emotions and instincts increase players' engagement to continue playing the game. The game designer and developer proposed the 6-11 framework, which contains 6 basic emotions and 11 instincts. For example, various serious games use an instinct to survive and thus to fight (eg, Re-Mission [38]) or to collect something (eg, VocabiCar [57] and Debugger 3.16: Hack'n'run [54]). Furthermore, LeBlanc's theory of 8 kinds of fun describes the desirable emotional responses evoked in players when they interact with the game [86].

Sense of Control

Players should feel in control over their actions in the game world [14]. In particular, players should have control and influence on the game world. For example, in the serious game PlayForward: Elm City Story, players can see how different actions lead to different outcomes [48]. Similarly, Escape from Diab allows players to influence the storyline and the characters [51,52]. As a result, the ability to influence the game world and in particular the story progress can motivate players to keep playing the game. Furthermore, serious games should support an optimal relationship between the player's actions and the game's reactions. For example, increased pedaling frequency in ErgoActive will always cause the character to rise [35].

Support Social Interactions

Bond and Beale [15] identified that good games offer some form of social interaction (see also the self-determination theory [92]). Social interactions in games are important for players who perform better depending on the game mode (eg, playing with friends or against them). For example, Pokémon GO [27] lets friends feed the player's creatures, and friends and family playing Dr Kawashima's Brain Training [49] can compete against each other. Vorderer et al [93] suggest that competitive elements are important for enjoyment. On the contrary, the work of Staiano et al [31] reveals that playing a cooperative version of the Nintendo Wii Active game is more effective than playing a competitive version. Thus, different game modes motivate players more or less. Especially if the players perform better in groups, collaborative and competitive multiplayer settings can contribute to motivating players [3]. For specific players, playing in a group (ie, multiplayer games) is more motivating than playing alone (ie, single-player games) [86]. Depending on the game purpose and target group, the game developers and designers should try to include different game modes so that a serious game is enjoyable for a broad player base.

Ensure Immersive Experience

Immersion in virtual environments can be increased by stimulating different human senses, especially by including appropriate audiovisual elements in the game. The game should use visceral, audio, and visual content to immerse players more deeply in the game [14]. Slater and Wilbur [94] describe immersion as the extent to which the computer system can deliver an illusion of a virtual environment to players. Thus, an immersive virtual environment should accommodate a wide range of appropriate synchronized sensory modalities. Recent studies already provide evidence that multimodal sensory stimulation improves the sense of presence (ie, the sense of "being there") and immersion [95]. For example, SnowWorld [50] successfully distracts players during rehabilitation (wound treatment) by immersing them in a virtual world. For a fully immersive virtual reality experience, serious games should include visual (eg, current-generation head-mounted displays), audio (eg, noise-canceling headphones), and haptic (eg, data gloves with force feedback or vibrations) feedback, as well as sense of smell (eg, smell dispenser) [96].

Media Presentation

One of the most apparent factors for immersive serious games is that they should have visually appealing graphics and appropriate sound effects. In particular, audiovisual elements in the game seize the attention of players [78].

Attractive Graphics

The included graphics should look attractive and engaging, as well as appropriate for the game purpose, application area, and the target group. For example, a game should have different designs for children, adults, and people with disabilities. The game designer should ensure clear interfaces without unnecessary information. Ravyse et al [78] furthermore suggest creating games that are high in realism; however, it should not be overloaded with unnecessary objects so that players do not get distracted from a specific task. For effective training, particularly for firefighters [97] or police training [59,98], the simulations should provide realistic virtual environments. However, in contrast to high-end graphics, reduced graphics can also be appropriate for some game types. For example, Minecraft (Mojang Studios) [99] is one of the most successful video games and it contains a world of blocks.

Appropriate Sound

Serious games should not only be visually appealing but should also include appropriate background music and sound effects. Previous studies have shown that audio influences the sense of presence, particularly in immersive virtual reality applications [100,101]. Martin-Niedecken et al [3] have expressed a similar view. The researchers show that music increases the motivation and immersion of the test subjects while playing the ExerCube. However, due to the players' varying music preferences, the choice of music could also be an important factor in motivation [102]. For example, the game Beat Saber allows players to create levels with custom songs [33].

Balance Between Serious and Game Part

The serious part and the game part of the game should be integrated and strongly connected, as seen in Table 3.

Table 3. Summary of quality criteria for balance between the serious and game part.

Quality criteria for balance and relevant quality aspects	Explanation
Integrated serious part with gameplay	
Embedding characterizing goal into the gameplay	• Integrate the characterizing goal into the gameplay • Learning/training tasks must be related to the game and should be connected to the game elements
Scientific foundation	• Include interdisciplinary teams; game designer and domain experts should work together (also together with the target group) • Include state of the art in the relevant disciplines
Interaction technology	
Appropriate interaction technology	• Interaction technology must be suitable for the target group (ie, their physical and mental ability and game purpose)
Intuitive game mechanics and natural mapping	• Provide tutorials for complex games; otherwise, players should discover the game mechanics themselves • Intuitive use of game controls (eg, the *WASD* keys to move and space bar to jump) • Enable natural mapping between technology and gameplay
No simplifying of the learning and/or training process due to technical features	• Interaction technology must support players in achieving the characterizing goal • Ensure accurate tracking to prevent cheating in exergames
Avoid adverse effects	• Low risk of accidents, injuries, or overload • Avoid technical issues and ensure easy maintenance

Integrated Serious Part With Gameplay

Serious games should embed the characterizing goal into the gameplay and the characterizing goal should not be avoidable.

Embedding Characterizing Goal Into the Gameplay

The gameplay experience includes, among other factors, an imaginative immersion (ie, immersion in the game world and the story) [95]. Thus, to motivate players, learning/training tasks need to be embedded into the immersive gameplay, such as in a story or narrative. Learning and training tasks should be directly related to the game and should be connected with the game elements and environment. For example, the educational game Addy (Coktel Vision) does not integrate learning into the game mechanics but uses a game only as a reward for learning. In this case, the serious part is not integrated into the game.

High-quality serious games should always integrate the serious part with the gameplay. For example, in adventure-based games, such as Meister Cody – Namagi and Meister Cody – Talasia, the tasks are embedded in a narrative and the story only proceeds when the player solves a problem. Similarly, Semideus [62] integrates the player's knowledge of rational numbers seamlessly into the gameplay. This close connection between the learning/training tasks and the storytelling makes the game even more motivating for children. An engaging story can motivate players; however, games without any story can still be very successful (eg, ExerCube [34]). Therefore, a story does not necessarily have to be profound or fascinating, but depending on the type of game, it can still be motivating.

Scientific Foundation

To develop an effective and attractive serious game, members of an interdisciplinary team of game designers, programmers, artists, and domain experts have to work together throughout the entire development process. As proposed by Martin-Niedecken et al [3], the target group should also be involved in the design process from the beginning (ie, participatory game design). This interdisciplinary team has to establish a balance between the disciplinary standards and requirements and the interdisciplinary integration under the twofold mission of serious games. Thus, for a high-quality serious game, one needs scientific foundations on both sides [1]. However, due to the limited development budget for serious games, there are often no resources for hiring professional game designers or artists. Thus, the challenge of an optimal balance between the serious and the game parts remains, making it difficult for the teams to develop a high-quality serious game that is entertaining and also fulfills its characterizing goal. Nevertheless, teams should continue to ensure an appropriate balance.

Interaction Technology

Depending on the serious game and the target group, adequate interaction technology must be deliberately chosen based on the specific needs of users and the game's purpose.

Appropriate Interaction Technology

The interaction technology must be suitable for the physical and mental ability of the target group. For example, the Kinect sensor (Microsoft Corp) and the Leap Motion Controller (Ultraleap) are particularly suitable for rehabilitation games, such as games for players with Parkinson disease or players who have had strokes, since these players often cannot hold or wear additional sensors [74]. Thus, depending on the target group, serious games must be presented by appropriate technology, including the appropriate visual display, speakers, and haptic interfaces, as well as suitable game-specific controllers (eg, gamepads or joysticks) [103]. The usage of an innovative technology should be appropriate for the game purpose as well; for instance, a head-mounted display can be annoying during sports, especially if the player is excessively sweating. However, commercial virtual reality games such as Beat Saber [33] have already proven that immersive environments are engaging and motivating despite sweating.

Intuitive Game Mechanics and Natural Mapping

It should be easy for players to understand how to play. Otherwise, the game should provide a tutorial. A tutorial should introduce not only game mechanics (how to do something) but also the gameplay (what to do). For inexperienced players, the controls should be basic enough to learn quickly, whereas experienced players can use advanced options. Desurvire and Wiberg [104] proposed principles for game designers to create better tutorials. In general, the game should be developed in such a way that players do not need to read a tutorial in order to play [14]. The work of Andersen et al [105] shows that tutorials are only appropriate for complex games and that simpler games should allow the players to discover the game mechanics themselves. Regardless of which interaction technology is used, the game controls should be intuitive, such as using the *WASD* keys to move and the space bar to jump (well-known game controls).

In addition to intuitive game mechanics, a game should also ensure natural mapping between technology and gameplay; in other words, the game should naturally map its controls to changes in the virtual environment [14]. For example, games for health should sense the movements of players to trigger the corresponding game responses [103]. Beat Saber [33] fulfills this requirement and allows players to intuitively move in the real world by simply walking around and using arm movements to slash the blocks.

No Simplifying of the Learning/Training Process Due to Technical Features

The interaction technology must support players in achieving the characterizing goal and must not impair or even disturb the learning/training processes. In exergames, players should not have advantages due to poor interaction technology (eg, cheating in Nintendo Wii sports games). The system must ensure that players are moving their bodies as required by the game [25]. Without accurate tracking, Wii players can sit on a couch and successfully play the game with a controller without performing the desired physical exercise. In this case, the system can be tricked and players only pretend to carry out a movement correctly. A study by Marks et al [106] shows that games played with Wii controllers require less physical activity than games played with the Kinect. To address this, exergames should track and detect motions accurately and in real time to ensure that players correctly execute all exercises. If the movement of a specific body part is required, this body part should be explicitly tracked to avoid cheating. For example, the ExerCube uses virtual reality trackers (HTC Vive; HTC Corp) to track upper body movements accurately [3].

Avoid Adverse Effects

Serious games should ensure that no accidents or injuries can happen, as older people can easily stumble and fall. Especially in exergames, the movements have to be tracked accurately so that players perform the exercises correctly and do not make mistakes that cause injuries. If required, the technology should be adapted and personalized to the specific target groups. For instance, the game BalanceFit uses a stability frame for secure balance training so that older people with heterogeneous skills are able to play the game [35]. Furthermore, physical or mental overload should be prevented by adequate monitoring of the player's psychophysical state or through regular breaks suggested by the game.

Independent of the application area, all games should avoid technical issues and should be easy to maintain [15]. Moreover, in immersive virtual reality, it is important to maintain a high frame rate, low latencies, and fast synchronization to avoid cybersickness. Cybersickness is caused by perceived motion or sensory mismatch, usually reported in a virtual roller coaster or car simulations. The latency between a user's movement and visual feedback in virtual reality has a significant impact on user experience and performance [107].

Conclusion

In this paper, we proposed criteria for high-quality serious games. We examined various serious games and existing heuristics from game-related literature to specify quality criteria for effective and attractive serious games. The suggested quality criteria were furthermore discussed in workshops with domain experts. We introduced quality criteria for the serious and game aspects, as well as the balance between them. First, high-quality serious games should keep the characterizing goal in focus and should use appropriate methods for the specific application area and target group. Serious games should provide suitable feedback so that players can assess their progress and work towards achieving the characterizing goal. The effectiveness of serious games should be proven in scientific studies or by winning game awards. Second, high-quality serious games should be fun and enjoyable. They must ensure player engagement and should keep the players in flow (ability vs skills). Finally, the double mission of serious games, that is, the balance between the serious and the game part, must be ensured. Therefore, high-quality serious games should embed the characterizing goal into the gameplay so that engaging in the serious part is mandatory for playing the game. Furthermore, the interaction technology should be suitable for the target group and game purpose.

In future work, we want to evaluate the proposed quality criteria, weight them, and determine a score to specify high-quality serious games, thereby proposing a quality mark. We hope that the proposed criteria will encourage game designers, developers, and researchers in the future to develop more high-quality serious games.

Acknowledgments

This work has been funded by the European Union (IWB-EFRE-Program Hessen) and the Hessian Ministry of Economics, Energy, Transport and Housing (Digitales Hessen) as part of the Wissens- und Technologietransfer Serious Games project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Classification of serious games for health.
[PDF File (Adobe PDF File), 83 KB - games_v8i3e19037_app1.pdf]

Multimedia Appendix 2
Classification of educational games.
[PDF File (Adobe PDF File), 41 KB - games_v8i3e19037_app2.pdf]

References

1. Dörner R, Göbel S, Effelsberg W, Wiemeyer J. Serious Games: Foundations, Concepts and Practice. Cham, Switzerland: Springer International Publishing; 2016.
2. Kuhn J, Holling H. Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. Adv Cogn Psychol 2014;10(2):59-67 [FREE Full text] [doi: 10.5709/acp-0157-2] [Medline: 25157301]
3. Martin-Niedecken A, Rogers K, Vidal L, Mekler E, Segura E. ExerCube vs personal trainer: valuating a holistic, immersive, and adaptive fitness game setup. 2019 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, United Kingdom p. 1-15. [doi: 10.1145/3290605.3300318]
4. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
5. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
6. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008 Aug;122(2):e305-e317 [FREE Full text] [doi: 10.1542/peds.2007-3134] [Medline: 18676516]
7. Beale IL, Kato PM, Marin-Bowling VM, Guthrie N, Cole SW. Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. J Adolesc Health 2007 Sep;41(3):263-270. [doi: 10.1016/j.jadohealth.2007.04.006] [Medline: 17707296]
8. Trash Monster. Bunny & Gnome. 2016. URL: https://muell-ag.de/
9. Orwell: Keeping an Eye on You. Osmotic Studios. 2016. URL: https://www.osmoticstudios.com/orwell-keeping-an-eye-on-you/
10. Civilization. Take-Two Interactive Software, Inc. 2016. URL: https://civilization.com/
11. Age of Empires. Microsoft. 1997. URL: https://www.ageofempires.com
12. Assassin's Creed. Ubisoft Entertainment. 2018. URL: https://assassinscreed.ubisoft.com [accessed 2020-05-01]
13. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Comput Entertain 2005 Jul 01;3(3):3. [doi: 10.1145/1077246.1077253]
14. Desurvire H, Wiberg C. Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. In: Lecture Notes in Computer Science. Berlin, Germany: Springer Berlin Heidelberg; 2009 Presented at: International Conference on Online Communities and Social Computing; July 19-24, 2009; San Diego, CA p. 557-566. [doi: 10.1007/978-3-642-02774-1_60]

15. Bond M, Beale R. What makes a good game? Using reviews to inform design. 2009 Presented at: People and Computers XXIII Celebrating People and Technology; Sep 1-5, 2009; Churchill College, Cambridge, UK p. 418-422. [doi: 10.14236/ewic/hci2009.52]
16. Göbel S, Gutjahr M, Steinmetz R. What makes a good serious game-conceptual approach towards a metadata format for the descriptionevaluation of serious games. 2011 Presented at: 5th European Conference on Games Based Learning; Oct 20-21, 2011; Athens, Greece p. 202-210.
17. Annetta LA. The “I's” Have It: A Framework for Serious Educational Game Design. Review of General Psychology 2010;14(2):105-112. [doi: 10.1037/a0018985]
18. Mueller F, Isbister K. Movement-based game guidelines. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 28-May 1, 2014; Toronto, Ontario, Canada p. 2191-2200. [doi: 10.1145/2556288.2557163]
19. Göbel S, Caserman P, Hansen J, et al. DIN SPEC 91380. Serious Games Metadata Format 2018. [doi: 10.31030/2853739]
20. Rego P, Moreira PM, Reis LP. Serious games for rehabilitation: A survey and a classification towards a taxonomy. 2010 Presented at: 5th Iberian Conference on Information Systems and Technologies; June 16-19, 2010; Santiago de Compostela, Spain p. 1-6.
21. Krathwohl D, Anderson L. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. London, England: Longman; 2009.
22. Carvalho MB, Bellotti F, Berta R, De Gloria A, Sedano CI, Hauge JB, et al. An activity theory-based model for serious games analysis and conceptual design. Computers & Education 2015 Sep;87:166-181. [doi: 10.1016/j.compedu.2015.03.023]
23. Djaouti D, Alvarez J, Jessel J. Classifying serious games: The G/P/S model. In: Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Hershey, PA: IGI Global; 2011:118-136.
24. Sawyer B, Smith P. Serious games taxonomy. 2008 Presented at: Serious Games Summit at the Game Developers Conference; Feb 18-22, 2008; San Francisco, CA. [doi: 10.1145/973801.973811]
25. Laamarti F, Eid M, El Saddik A. An Overview of Serious Games. International Journal of Computer Games Technology 2014;2014:1-15. [doi: 10.1155/2014/358152]
26. Serious Games Information Center. 2018. URL: https://seriousgames-portal.org [accessed 2020-05-01]
27. Pokémon GO. Niantic, Inc. 2016. URL: https://pokemongolive.com/ [accessed 2020-05-01]
28. Dance Dance Revolution. Konami. 2010. URL: https://www.konami.com/games/asia/en/products/ddr_a/ [accessed 2020-05-01]
29. Lanningham-Foster L, Jensen TB, Foster RC, Redmond AB, Walker BA, Heinz D, et al. Energy expenditure of sedentary screen time compared with active screen time for children. Pediatrics 2006 Dec;118(6):e1831-e1835. [doi: 10.1542/peds.2006-1087] [Medline: 17142504]
30. Murphy EC, Carson L, Neal W, Baylis C, Donley D, Yeater R. Effects of an exercise intervention using Dance Dance Revolution on endothelial function and other risk factors in overweight children. Int J Pediatr Obes 2009;4(4):205-214. [doi: 10.3109/17477160902846187] [Medline: 19922034]
31. Staiano AE, Abraham AA, Calvert SL. Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention. Obesity (Silver Spring) 2013 Mar;21(3):598-601 [FREE Full text] [doi: 10.1002/oby.20282] [Medline: 23592669]
32. Sohnsmeyer J, Gilbrich H, Weisser B. Effect of a six-week-intervention with an activity-promoting video game on isometric muscle strength in elderly subjects. International Journal of Computer Science in Sport (International Association of Computer Science in Sport) 2010;9(2):75-79 [FREE Full text]
33. Beat Saber. Beat Games. 2018. URL: https://beatsaber.com [accessed 2020-05-01]
34. The ExerCube. Sphery Ltd. 2018. URL: https://sphery.ch/der-exercube/ [accessed 2020-05-01]
35. Hardy S, Dutz T, Wiemeyer J, Göbel S, Steinmetz R. Framework for personalized and adaptive game-based training programs in health sport. Multimed Tools Appl 2014 May 21;74(14):5289-5311. [doi: 10.1007/s11042-014-2009-z]
36. Hardy S, Göbel S, Steinmetz R. Adaptable and personalized game-based training system for fall prevention. 2013 Presented at: 21st ACM international conference on Multimedia; Oct 21-25, 2013; Barcelona, Spain p. 431-432. [doi: 10.1145/2502081.2502255]
37. Hoffmann K, Hardy S, Wiemeyer J, Göbel S. Personalized Adaptive Control of Training Load in Cardio-Exergames-A Feasibility Study. Games for Health Journal 2015 Dec;4(6):470-479. [doi: 10.1089/g4h.2014.0073] [Medline: 26484388]
38. Re-Mission. Hopelab. 2006. URL: https://www.re-mission2.org [accessed 2020-05-01]
39. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
40. Fiellin LE, Hieftje KD, Pendergrass TM, Kyriakides TC, Duncan LR, Dziura JD, et al. Video Game Intervention for Sexual Risk Reduction in Minority Adolescents: Randomized Controlled Trial. J Med Internet Res 2017 Sep 18;19(9):e314 [FREE Full text] [doi: 10.2196/jmir.8148] [Medline: 28923788]
41. Virtual Reality Pain Reduction. College of Engineering, University of Washington. URL: http://www.vrpain.com [accessed 2020-05-01]

42. Nouchi R, Taki Y, Takeuchi H, Hashizume H, Akitsuki Y, Shigemune Y, et al. Brain training game improves executive functions and processing speed in the elderly: a randomized controlled trial. PLoS One 2012 Jan;7(1):e29676 [FREE Full text] [doi: 10.1371/journal.pone.0029676] [Medline: 22253758]
43. Nouchi R, Taki Y, Takeuchi H, Hashizume H, Nozawa T, Kambara T, et al. Brain training game boosts executive functions, working memory and processing speed in the young adults: a randomized controlled trial. PLoS One 2013;8(2):e55518 [FREE Full text] [doi: 10.1371/journal.pone.0055518] [Medline: 23405164]
44. Pokémon GO. Niantic, Incn. Google Play Store. 2016. URL: https://play.google.com/store/apps/details?id=com.nianticlabs.pokemongo [accessed 2020-05-01]
45. Gough C. Pokémon Go player spending worldwide from 2016 to 2019. Statista. 2020 Mar 5. URL: https://www.statista.com/statistics/882474/pokemon-go-all-time-player-spending-countries/ [accessed 2020-05-01]
46. Best of 2019 virtual reality: The top VR-only experiences as measured by gross revenue this year. Steam. 2019. URL: https://store.steampowered.com/sale/2019_top_vr [accessed 2020-05-01]
47. Beat Saber. Steam. 2018. URL: https://store.steampowered.com/app/620980/Beat_Saber/ [accessed 2020-05-01]
48. PlayForward: Elm City Stories. Schell Games. 2013. URL: https://www.schellgames.com/games/playforward/ [accessed 2020-05-01]
49. Dr. Kawashima’s Brain Training (Brain Age). Nintendo. 2005. URL: https://www.nintendo.com/games/detail/brain-age-train-your-brain-in-minutes-a-day-wii-u/ [accessed 2020-05-01]
50. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
51. Thompson D, Baranowski T, Buday R, Baranowski J, Thompson V, Jago R, et al. Serious Video Games for Health How Behavioral Science Guided the Development of a Serious Video Game. Simul Gaming 2010 Aug 01;41(4):587-606 [FREE Full text] [doi: 10.1177/1046878108328087] [Medline: 20711522]
52. Escape from Diab and Nanoswarm: Invasion from inner space. Archimage, Inc. 2006. URL: http://archimage.com [accessed 2020-06-19]
53. Görgen R, Huemer S, Schulte-Körne G, Moll K. Evaluation of a digital game-based reading training for German children with reading disorder. Computers & Education 2020 Jun;150:103834. [doi: 10.1016/j.compedu.2020.103834]
54. Debugger 3.16 Hack'n'Run. Spiderwork Games. 2017. URL: http://www.spiderwork-games.com/debugger.htm [accessed 2020-05-01]
55. Ninaus M, Kiili K, McMullen J, Moeller K. Assessing fraction knowledge by a digital game. Computers in Human Behavior 2017 May;70:197-206. [doi: 10.1016/j.chb.2017.01.004]
56. Kiili K, Ketamo H. Evaluating Cognitive and Affective Outcomes of a Digital Game-Based Math Test. IEEE Trans. Learning Technol 2018 Apr 1;11(2):255-263. [doi: 10.1109/tlt.2017.2687458]
57. VocabiCar. Quantumfrog. 2017. URL: https://vocabicar.de [accessed 2020-05-01]
58. Orwell: Ignorance is Strength. Osmotic Studios. 2018. URL: https://www.osmoticstudios.com/orwell-ignorance-is-strength/ [accessed 2020-05-01]
59. Bertram J, Moskaliuk J, Cress U. Virtual training: Making reality work? Computers in Human Behavior 2015 Feb;43:284-292. [doi: 10.1016/j.chb.2014.10.032]
60. Learning game for children struggling in math and reading. Meister Cody. 2013. URL: https://www.meistercody.com [accessed 2020-05-01]
61. Debugger 3.16: Hack'n'Run. Steam. 2017. URL: https://store.steampowered.com/app/339720/Debugger_316_HacknRun/ [accessed 2020-05-01]
62. Semideus. Flow Factory Oy. 2016. URL: http://www.flowfactory.fi/semideus/ [accessed 2020-05-01]
63. Orwell: Keeping an Eye On You. Steam. 2016. URL: https://store.steampowered.com/app/491950/Orwell_Keeping_an_Eye_On_You/ [accessed 2020-05-01]
64. Orwell: Ignorance is Strength. Steam. 2018. URL: https://store.steampowered.com/app/633060/Orwell_Ignorance_is_Strength/ [accessed 2020-05-01]
65. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
66. Schiesel S. Classes turn to video game that works legs. The New York Times 2007 Apr 30 [FREE Full text]
67. Peng W, Crouse JC, Lin J. Using active video games for physical activity promotion: a systematic review of the current state of research. Health Educ Behav 2013 Apr;40(2):171-192. [doi: 10.1177/1090198112444956] [Medline: 22773597]
68. Gough C. All-time best-selling console games based on global unit sales as of April 2020. Statista. 2020 Apr 1. URL: https://www.statista.com/statistics/264530/all-time-best-selling-console-games-worldwide/ [accessed 2020-04-01]
69. Bartolomé NA, Zorrilla A, Zapirain B. Can game-based therapies be trusted? Is game-based education effective? A systematic review of the Serious Games for health and education. 2011 Presented at: 16th International Conference on Computer Games (CGAMES); July 27-30, 2011; Louisville, KY p. 275-282. [doi: 10.1109/cgames.2011.6000353]

70. 2019 Essential Facts About the Computer and Video Game Industry. Entertainment Software Association. 2019. URL: https://www.theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/ [accessed 2020-05-01]
71. Clement J. Most popular Google Play app categories 2020. Statista. 2020 May 5. URL: https://www.statista.com/statistics/279286/google-play-android-app-categories/ [accessed 2020-05-01]
72. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 2013;105(2):249-265. [doi: 10.1037/a0031311]
73. Lü. Lü Interactive Playground. 2017. URL: https://www.play-lu.com [accessed 2020-05-01]
74. Garcia-Agundez A, Folkerts A, Konrad R, Caserman P, Tregel T, Goosses M, et al. Recent advances in rehabilitation for Parkinson's Disease with Exergames: A Systematic Review. J Neuroeng Rehabil 2019 Jan 29;16(1):1-17 [FREE Full text] [doi: 10.1186/s12984-019-0492-1] [Medline: 30696453]
75. Garcia-Agundez A, Folkerts A, Konrad R, Caserman P, Göbel S, Kalbe E. PDDanceCity: An Exergame for Patients with Idiopathic Parkinson’s Disease and Cognitive Impairment. 2017 Presented at: Mensch und Computer; Sep 10-13, 2017; Regensburg, Germany p. 381-386. [doi: 10.18420/muc2017-mci-0334]
76. Giessen HW. Serious Games Effects: An Overview. Procedia - Social and Behavioral Sciences 2015 Feb;174:2240-2244. [doi: 10.1016/j.sbspro.2015.01.881]
77. Whyte EM, Smyth JM, Scherf KS. Designing Serious Game Interventions for Individuals with Autism. J Autism Dev Disord 2015 Dec;45(12):3820-3831. [doi: 10.1007/s10803-014-2333-1] [Medline: 25488121]
78. Ravyse WS, Seugnet Blignaut A, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virtual Reality 2016 Sep 20;21(1):31-58. [doi: 10.1007/s10055-016-0298-4]
79. Graham K, Schofield D. Rock god or game guru: Using Rocksmith to learn to play a guitar. Journal of Music, Technology & Education 2018 Aug 01;11(1):65-82. [doi: 10.1386/jmte.11.1.65_1]
80. Ryan RM, Weinstein N. Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. Theory and Research in Education 2009 Jun 25;7(2):224-233. [doi: 10.1177/1477878509104327]
81. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
82. Koster R. Theory of fun for game design. Sepastopol, CA: O'Reilly Media Inc; 2013.
83. Czikszentmihalyi M. Flow: The psychology of optimal experience. New York, NY: Harper & Row; 1990.
84. Crockett L. Easy to learn, difficult to master: Accessible front-ends to challenging science assignments. 2016 Presented at: 10th European Conference on Games Based Learning: ECGBL 2016; Oct 6-7, 2016; University of the West, Paisley, Scotland p. 144-151.
85. Dillon R. On the Way to Fun: An Emotion-Based Approach to Successful Game Design. Natick, MA: A K Peters, Ltd; 2010.
86. Hunicke R, LeBlanc M, Zubek R. MDA: A formal approach to game design and game research. : Proceedings of the AAAI Workshop on Challenges in Game AI; 2004 Presented at: AAAI Workshop on Challenges in Game AI; July 25-29, 2004; San Jose, CA p. 1-5.
87. Calvillo-Gámez E, Cairns P, Cox A. Assessing the core elements of the gaming experience. In: Game User Experience Evaluation. Cham, Switzerland: Sprinter International Publishing; 2015:37-62.
88. Bartle R. Hearts, clubs, diamonds, spades: Players who suit MUDs. Journal of MUD Research 1996;1(1):19 [FREE Full text]
89. Scoresby J, Shelton BE. Visual perspectives within educational computer games: effects on presence and flow within virtual immersive learning environments. Instr Sci 2010 Feb 4;39(3):227-254. [doi: 10.1007/s11251-010-9126-5]
90. Pivec M. Editorial: Play and learn: potentials of game-based learning. Br J Educ Technol 2007 May;38(3):387-393. [doi: 10.1111/j.1467-8535.2007.00722.x]
91. Whitehead A, Johnston H, Nixon N, Welch J. Exergame effectiveness: what the numbers can tell us. New York, NY, United States: Association for Computing Machinery; 2010 Presented at: 5th ACM SIGGRAPH Symposium on Video Games; July 28-29, 2010; Los Angeles, CA p. 55-62. [doi: 10.1145/1836135.1836144]
92. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
93. Vorderer P, Hartmann T, Klimmt C. Explaining the enjoyment of playing video games: the role of competition. 2003 Presented at: Second International Conference on Entertainment Computing; May 8-10, 2003; Pittsburgh, PA p. 1-9.
94. Slater M, Wilbur S. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. Presence: Teleoperators and Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
95. Ermi L, Mäyrä F. Fundamental components of the gameplay experience: Analysing immersion. Worlds in play: International perspectives on digital games research 2005;37(2):37-53 [FREE Full text]
96. Caserman P, Garcia-Agundez A, Goebel S. A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications. IEEE Trans. Visual. Comput. Graphics 2019:1-20. [doi: 10.1109/tvcg.2019.2912607]

97. Eller C, Bittner T, Dombois M, Rüppel U. Collaborative immersive planning and training scenarios in VR. 2018 Presented at: Workshop of the European Group for Intelligent Computing in Engineering (EG-ICE); June 10-13, 2018; Lausanne, Switzerland p. 164-185. [doi: 10.1007/978-3-319-91635-4_9]
98. Caserman P, Zhang H, Zinnäcker J, Göbel S. Development of a directed teleport function for immersive training in virtual reality. 2019 Presented at: International Conference on Virtual Worlds and Games for Serious Applications (VS-Games); Sep 4-6, 2019; Vienna, Austria p. 1-8. [doi: 10.1109/vs-games.2019.8864599]
99. Minecraft. Mojang. 2009. URL: https://www.minecraft.net/ [accessed 2020-05-01]
100. Cummings JJ, Bailenson JN. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. Media Psychology 2015 May 15;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
101. Kern AC, Ellermeier W. Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence. Front. Robot. AI 2020 Feb 21;7(20). [doi: 10.3389/frobt.2020.00020]
102. Sanders T, Cairns P. Time perception, immersion and music in videogames. 2010 Presented at: 24th BCS Conference on Human Computer Interaction; Sep 6-10, 2010; Dundee, UK p. 160-167. [doi: 10.14236/ewic/hci2010.21]
103. Wiemeyer J. Towards a generic framework for serious games. 2019 Presented at: 12th International Symposium on Computer Science in Sport (IACSS); July 8-10, 2019; Moscow, Russia p. 193-200. [doi: 10.1007/978-3-030-35048-2_23]
104. Desurvire H, Wiberg C. Game User Experience Evaluation. In: User Experience Design for Inexperienced Gamers: GAP—Game Approachability Principles. Cham, Switzerland: Springer; 2015:169-186.
105. Andersen E, O'Rourke E, Liu YE, Snider R, Lowdermilk J, Truong D, et al. The impact of tutorials on games of varying complexity. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 59-68. [doi: 10.1145/2207676.2207687]
106. Marks DW, Rispen L, Calara G. Greater physiological responses while playing XBox Kinect compared to Nintendo Wii. International Journal of Exercise Science 2015;8(2):7 [FREE Full text] [doi: 10.13140/RG.2.1.3302.5446]
107. Caserman P, Martinussen M, Göbel S. Effects of end-to-end latency on user experience and performance in immersive virtual reality applications. 2019 Presented at: Joint International Conference on Entertainment Computing and Serious Games; Nov 11-15, 2019; Arequipa, Peru p. 57-69. [doi: 10.1007/978-3-030-34644-7_5]

Abbreviations

SG-MDF: DIN SPEC 91380 Serious Games Metadata Format

Edited by M Birk; submitted 03.04.20; peer-reviewed by M Kickmeier-Rust, H Söbke; comments to author 02.05.20; revised version received 14.05.20; accepted 29.05.20; published 24.07.20.

Please cite as:
Caserman P, Hoffmann K, Müller P, Schaub M, Straßburg K, Wiemeyer J, Bruder R, Göbel S
Quality Criteria for Serious Games: Serious Part, Game Part, and Balance
JMIR Serious Games 2020;8(3):e19037
URL: http://games.jmir.org/2020/3/e19037/
doi:10.2196/19037
PMID:32706669

Original Paper

Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization

Claudia Jenny[1], Dr; Christoph Reuter[1], Prof Dr

[1]Musicological Department, University of Vienna, Vienna, Austria

Corresponding Author:
Christoph Reuter, Prof Dr
Musicological Department
University of Vienna
Spitalgasse 2, Hof 9
Vienna, 1090
Austria
Phone: 43 1 4277 44280
Fax: 43 1 4277 9442
Email: christoph.reuter@univie.ac.at

Abstract

Background: In order to present virtual sound sources via headphones spatially, head-related transfer functions (HRTFs) can be applied to audio signals. In this so-called binaural virtual acoustics, the spatial perception may be degraded if the HRTFs deviate from the true HRTFs of the listener.

Objective: In this study, participants wearing virtual reality (VR) headsets performed a listening test on the 3D audio perception of virtual audiovisual scenes, thus enabling us to investigate the necessity and influence of the individualization of HRTFs. Two hypotheses were investigated: first, general HRTFs lead to limitations of 3D audio perception in VR and second, the localization model for stationary localization errors is transferable to nonindividualized HRTFs in more complex environments such as VR.

Methods: For the evaluation, 39 subjects rated individualized and nonindividualized HRTFs in an audiovisual virtual scene on the basis of 5 perceptual qualities: localizability, front-back position, externalization, tone color, and realism. The VR listening experiment consisted of 2 tests: in the first test, subjects evaluated their own and the general HRTF from the Massachusetts Institute of Technology Knowles Electronics Manikin for Acoustic Research database and in the second test, their own and 2 other nonindividualized HRTFs from the Acoustics Research Institute HRTF database. For the experiment, 2 subject-specific, nonindividualized HRTFs with a minimal and maximal localization error deviation were selected according to the localization model in sagittal planes.

Results: With the Wilcoxon signed-rank test for the first test, analysis of variance for the second test, and a sample size of 78, the results were significant in all perceptual qualities, except for the front-back position between own and minimal deviant nonindividualized HRTF (P=.06).

Conclusions: Both hypotheses have been accepted. Sounds filtered by individualized HRTFs are considered easier to localize, easier to externalize, more natural in timbre, and thus more realistic compared to sounds filtered by nonindividualized HRTFs.

(JMIR Serious Games 2020;8(3):e17576) doi:10.2196/17576

KEYWORDS

head-related transfer function; sound localization; immersive virtual reality; binaural virtual acoustics; 3D audio perception

Introduction

Theories

The question raised in the article, "Binaural Technique: Do We Need Individual Recordings?" by Møller et al [1], is one that many researchers and developers still ask themselves. The increasing access to advanced virtual and augmented reality technologies gives this topic a particular immediacy. There are different schools of thought as to whether it is important to have personalized head-related transfer functions (HRTFs) for a realistic reproduction of auditory scenes via headphones in virtual reality (VR). The ability to adapt nonindividualized HRTFs via training [2] or the given tolerance by adding distance perception [3], auralization [4,5], auditory motion [6], and cross/multimodal perception [7-10] as well as different recorded auditory stimuli (eg, noise bursts, speech, music) still brings up the question in the title of Møller et al [1]. However, it is well accepted that individualization has a significant effect in sagittal plane sound localization with static target position without visual stimulus [1,11,12]. There are several VR studies where the focus lies on immersive VR, which is helpful in health care (the more immersive the better) [13] or in spatial navigation memory assessment [14], though considerations about immersive audio are missing in these studies. Audio is often neglected in VR studies (eg, [15]), even regarding the sound quality [16], not to mention HRTFs. If and to what extent the perception quality of sound signals in VR can be improved by using individualized HRTFs has not been investigated yet.

Background

Acoustic localization is the ability to determine and report the position of a virtual sound source and is based on the processing of auditory localization features such as monaural and binaural features [17-21]. It is assumed that binaural and monaural spectral features are processed largely independently of each other [22,23]. While binaural disparities such as interaural time and level differences play an important role in sound localization in the lateral dimension (left-right), monaural spectral cues are known to determine the perceived position of the sound source in the sagittal planes (front-back and up-down). Sound localization in sagittal planes relies on spectral features caused by the filtering effects of the human body [24].

HRTFs describe the acoustic filter effect through the torso, head, and pinna [24-26]. A set of HRTFs (also called “binaural HRTF”) includes the primary localization cues: interaural time differences, interaural level differences, and the monaural spectral cues [24]. This acoustic filter of our own anatomy is individually different and highly frequency-dependent. When HRTFs are measured in the listener’s own ears, it is described as “individual,” “own,” or “listener-specific,” whereas “nonindividual,” “other,” or “generic” HRTF refer to measurements from a different listener, a dummy head, or a calculation from a model.

In order to present virtual sound sources via headphones, the audio signal can be filtered with HRTFs. In this so-called binaural virtual acoustics, the spatial perception may be limited if the used HRTFs deviate from the individualized HRTFs of the listener [11]. This can lead to incorrect virtual sound source positions or even to a localization within the head.

Prior Work

Individual features should be used to ensure realistic replication, as previous studies have shown that by using listener-specific HRTFs for headphone reproduction, subjects could locate the source of the sound just as accurately as if they were listening to something in free-field reproduction [27,28]. Their research results also showed that subjects with nonindividualized HRTFs have significantly greater localization errors, especially in the median plane, and that front-back confusion increases. However, the results of other studies show that subjects with nonindividualized HRTFs have no localization loss in the horizontal plane with voice stimuli [29] nor do inexperienced subjects acknowledge any significant impact on front-back confusion with individualized HRTFs [30]. Furthermore, studies [31-33] have shown a worsening of externalization or a significant increase in the inside-head localization and an increase in the localization errors in the distance perception in subjects who heard stimuli with nonindividualized HRTFs. Romigh and Simpson [34] confirmed that the replacement of listener-specific interaural features by generic interaural features did no harm but replacing listener-specific monaural features with generic monaural features did interfere with localization performance. Localization models such as the probabilistic model for stationary localization errors in sagittal planes [35] can be used to predict localization errors, which a listener would have had with HRTFs from another listener.

Goal of This Study

Our study examines the need for individualization of HRTFs in headphone reproduction and the impact of customizability of binaural performance in audiovisual virtual environments. The aim of the study was to find out if and to what extent the perception quality of sound signals in VR can be improved by using individualized HRTFs.

The hypotheses of this experiment can be summarized as follows:

Hypothesis 1: General HRTFs such as the KEMAR (Knowles Electronics Manikin for Acoustic Research) HRTF lead to limitations of 3D audio perception in VR.

Hypothesis 2: The localization model for stationary localization errors is transferable to nonindividualized HRTFs in a multimodal representation.

For the general HRTF, we have chosen the KEMAR HRTF from the Massachusetts Institute of Technology (MIT) KEMAR database [36], which is one of the most widely used HRTFs in both science and industry. The artificial head used to obtain the data has the dimensions of an average human ear and body. We assumed that generic HRTFs lead to limitations of 3D audio perception in VR, such as sound sources would be more difficult to localize and internalize and the tone color would be unnatural, and in general, perceived as unrealistic. However, front-back confusions would be unlikely because listeners were able to move their heads [37]. For the nonindividualized HRTFs, we have chosen HRTFs from the Acoustic Research Institute (ARI) HRTF database. The difference between the KEMAR HRTF and the HRTFs from the ARI database is that the KEMAR HRTF is measured from a dummy head and the HRTFs from the ARI database are measured from human subjects.

Methods

General Information

For the study, 39 subjects rated individualized and nonindividualized HRTFs in an audiovisual virtual scene by using a questionnaire, which consisted of 5 perceptual qualities (localizability, front-back position, externalization, tone color, realism; see definitions in Table 1) and was based on the spatial audio quality inventory [38] and the study of Simon et al [39]. A head-mounted display was used to present an acoustically located flying dynamic sound source (drone) in a winter landscape environment. Switching HRTFs took place via touch controllers enabled with the plugin [40]. The filter algorithms took the listener interaction into account in real time. The VR listening experiment consisted of 2 tests: in the first test, subjects rated their own versus a general HRTF (MIT KEMAR dummy head, [36]) and in the second test, their own versus 2 other nonindividualized HRTFs from the ARI HRTF database. As a basis for the selection of nonindividualized HRTFs, the localization model in the sagittal plane by Baumgartner et al [35] was used, which predicts localization errors. For the experiment, 2 listener-specific, nonindividualized HRTFs with a minimum and maximum localization error deviation were selected. For the selection, 140 HRTFs from the ARI database were chosen.

Table 1. Perceptual qualities for the assessment of the audiovisual scene derived from the studies on spatial audio quality inventory [38] and that of Simon et al [39].

Perceptual quality	Circumscription	Scale end label
Localizability	If localizability is low, the spatial extent and location of a sound source are difficult to estimate or they appear diffuse. If localizability is high, a sound source is clearly delimited. Low/high localizability is often associated with high/low perceived extent of a sound source [38].	More difficult to easier
Front-back position	Refers to the position of a sound source before or behind the listener only. Impression of a position difference of a sound source caused by “reflecting” its position on the frontal plane going through the listener [38].	Confused / not confused
Externalization	Describes the distinctness with which a sound source is perceived within or outside the head regardless of the distance. Terminologically often enclosed between the phenomena of in-head localization and out-of-head localization [38].	More internalized to more externalized
Tone color bright to dark	Timbral impression, which is determined by the ratio of high-frequency to low-frequency components [38].	Darker to brighter
Realism	Sounds seem to come from real sources located around you [39].	Nonrealistic to realistic

Subjects

A total of 39 subjects took part in the study. All of them (males, 26/39, 67%; females, 13/39, 33%) had absolute hearing thresholds within the 20-dB range of the average normal hearing population in the frequency range between 0.125 kHz and 12.5 kHz. The 39 subjects had a mean (SD) age of 30.03 (6.738) years (age range, 22-47 years), and about half were experienced listeners (low expertise, 19 subjects; high expertise, 20 subjects). The “low expertise” group included, for instance, lay listeners, who might have been music lovers but were not trained musicians. The “high expertise” group included experienced listeners such as trained musicians, “Tonmeister,” and sound engineers [41]. To determine the required number of subjects, we conducted an a priori power analysis with the software program G*Power (Heinrich Heine University Düsseldorf). A two-sided *t* test with Wilcoxon signed-rank test (one sample case) was assumed, which resulted in a total sample size of 35 with an expected mean effect size of d=0.5 for an α error of .05 and a test power of 1–β of 80%.

HRTF Measurement

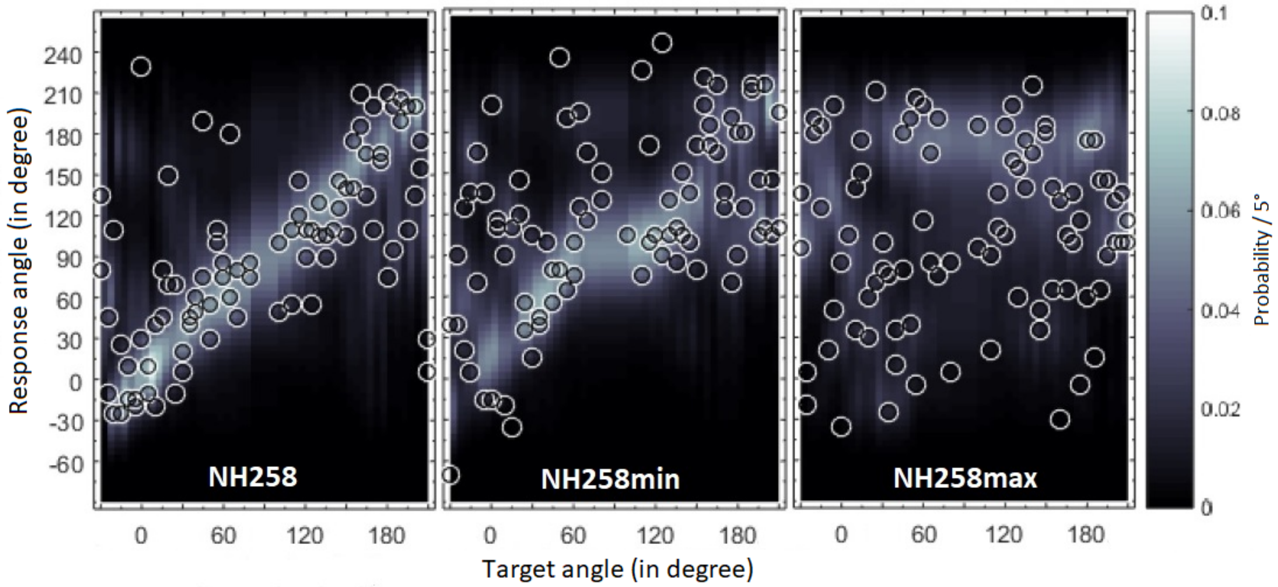
HTRFs were obtained for each subject individually by measuring in a semianechoic chamber. The same apparatus and procedure as reported by Majdak et al [42] were used. With a loudspeaker arc of 22 vertically arranged loudspeakers (custom-made with 10 BGS drivers, Vifa), 1550 measuring positions were achieved. The loudspeakers were arranged in the elevation direction from –30° to 80°, with a 5° spacing except between 70° and 80° with a 10° increment. The radius of the loudspeaker bow was 1.2 m. The rotation of the turntable with chair took place in 2.5° increments. For the recording in the ear canal, in-ear microphones (Sennheiser KE-4-211-2) were used. The microphones were connected to the digital audio interface via amplifiers (Radio Design Lamps FP-MP1). An electromagnetic tracking system (Flock of Birds, Ascension) was used to monitor the head position and orientation. As signals during the measurement were exponential, sine sweeps were used with a signal length of about 1.8 seconds, starting at 50 Hz and ending at 20 kHz. To reduce the total time taken to measure HRTFs, we used the multiple exponential sweep method [43]. The HRTF measurement procedure took approximately 60 minutes for each subject, including instruction, reference measurements, and adjustments. The measuring process itself took about 20 minutes.

Stimuli

As an acoustic stimulus, the synthetically generated stimulus, Gaussian white noise was selected. Gaussian white noise is often used as a stimulus in HRTF studies and was applied to simulate a drone noise. The stimulus was filtered by individualized and nonindividualized HRTFs. The selection of nonindividualized HRTFs in the second test was based on the sagittal plane localization model by Baumgartner et al [35]. This model can predict localization errors of static sound source positions in an auditory-only environment. For the experiment, 2 listener-specific, nonindividualized HRTFs with a minimum and maximum localization error deviation were selected. The deviations of the stationary localization errors were given by means of this model in quadrant error (QE) in percentage and root mean square local polar errors (PEs) in degrees. The QE and PE were calculated using the model via the Auditory Modeling Toolbox in Matlab [44]. Based on the subjects’ own HRTFs (template), it is possible to predict how large QE and

PE are when the subjects then hear another HRTF (target). For the range of minimum and maximum HRTFs, 2 conditions were defined to ensure comparability: 1st condition, minimum nonindividualized HRTF range, QE of 10%-30%, PE of 33°-42°, maximum nonindividualized HRTF range, QE of 30%-50%, and PE of 43°-52.25°; 2nd condition, minimum distance between individual/minimal and minimal/maximal HRTFs, QE of 3%, and PE of 3°. An individual, minimal, and a maximal deviant HRTF is shown in Figure 1 using the example of the listener NH258 (normal hearing listener number 258). The individualized HRTF of NH258 had an initial value of QE 18.2% and PE 36.1°. For example, a slightly minimal deviant HRTF would be that of NH157, with which NH258 would have QE of 21.3% and PE of 40.7°. A maximal deviant HRTF would be NH89, with which NH258 would have a QE of 40% and PE of 46.7°. For all HRTFs, the sensitivity parameter was set to the default value 0.5 in the model. All the selected minimal and maximal deviant HRTFs were calculated individually for each subject.

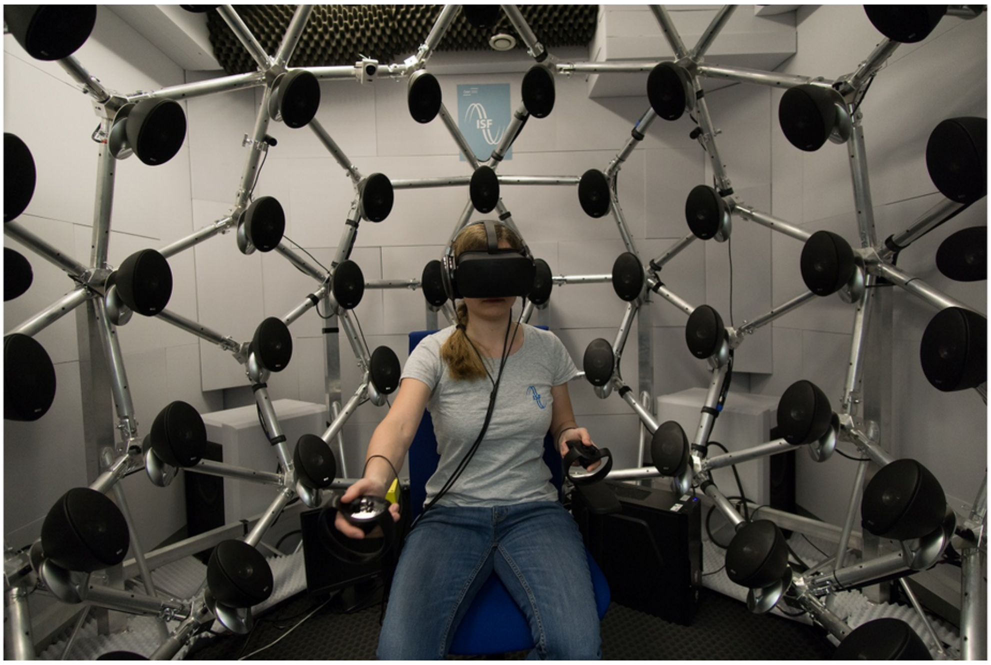
Figure 1. Localization model for prediction of localization errors in the sagittal plane. Probabilistic response predictions are encoded by brightness according to the color bar to the right. Predicted response angles are shown as open circles. NH258: normal hearing listener number 258; min: minimal deviant head-related transfer function; max: maximal deviant head-related transfer function.



Apparatus

The virtual acoustic stimuli were presented via headphones (HD 650, Sennheiser) in a semianechoic room. As shown in Figure 2, the listener was seated on a height-adjustable swivel chair in the middle of the room. The virtual visual environment, created with Unity version 2017.3.1f1, was presented via a head-mounted display (Oculus Rift CV1 headset, 2 PenTile organic laser-emitting diode displays, 2160×1200 combined resolution for both eyes, 90 Hz refresh rate, 110° FoV) including touch controller for switching the HRTFs using the Barebone gaming PC in Thermaltake housing with Intel (R) Core (TM) i5-6500 CPU 3.2 GHz processor, 16 GB RAM, 64-bit operating system (Windows 10), 200 GB SSD, GTX 1060 graphics card (6 GB VRAM), HDMI 1.3, 4x USB 3.0, 2x USB 2.0, mouse, keyboard, screen and 3 sensors for head-tracking. Stimuli were generated using the “SOFA (Spatially Oriented Format for Acoustics) Spatializer” plugin [40] and output with a 48-kHz sampling rate filtered with individualized and nonindividualized HRTFs. The “SOFA Spatializer” plugin is a Unity native plugin based on C/C++ for enabling playing HRTF in the SOFA format [45]. The virtual visual environment was created in C# by using the Unity game engine. Three tracking sensors captured the position and orientation of the head in real time. The front 2 sensors were connected with USB 3.0 and the back with USB 2.0. All sensors were fixed at the same height, slightly above the head height. The 2 front sensors had a distance of 1.8 m and the rear sensor had a straight line distance of 2.55 m from the farthest front sensor. The range of motion was 2.4×2.4 m². The sensors were 1.2 m away from the headset. The sensor settings thus corresponded to all specifications for the head-mounted display.

Figure 2. Experimental setup with the Oculus Rift head-mounted display in a semianechoic room. Test environment with open, dynamic, circumaural Sennheiser HD 650 headphones, 3 sensors for the tracking system, and touch controllers for switching the head-related transfer functions. The loudspeaker array was not in use in this experiment.



Description of the Test Environment

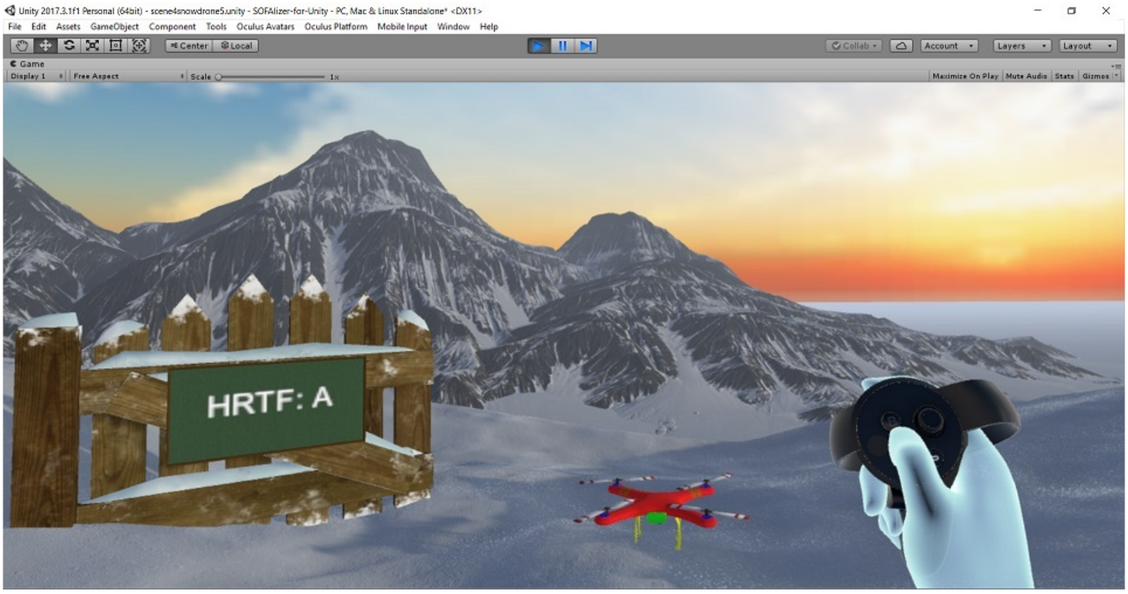
The VR listening test took about 60 minutes per subject. The subjects initially were each given an informed consent form and the list of attributes to familiarize themselves with the technical terms (Table 1). Before the actual VR listening test, a pilot study with 3 subjects was conducted to test if the experimental design worked. Two tests were performed.

Individualized and General HRTFs

In the first test, the subject was given a rather easy task: the evaluation of 2 HRTFs, that is, their own and the general KEMAR. Both the subject and the experimenter did not know which HRTF (A or B) was assigned (double-blind study). In addition, the set of questions concerning the HRTF was randomized in terms of the subject as well as the repetition. However, each subject was aware that one was the individualized HRTF and the other the general HRTF.

In the scene, a drone flew overhead along the sagittal planes, landing in front and flying back. In polar coordinates, the audiovisual stimulus flew between –30° and 210° with a distance of 1.2 m (same dimensions as HRTF measurement) back and forth. One animation cycle lasted 24 seconds: 10 seconds for each semicircle flight plus 2 seconds for each landing. A continuous Gaussian white noise was used as the auditory stimulus. The visual stimulus, the drone with 4 rotating rotor blades, served as a guide (visual aid) in which the position of the sound source was supposed to be straight. In addition, haptic touch was added. The avatar hands with the touch controllers were used to switch the HRTFs with a simultaneous display on a fence with a blackboard. Figure 3 shows a screenshot from the subject's point of view. At the start of the scene, the subject was given about 2 minutes to familiarize herself with the scene and the touch controllers. With the right touch controller, the subject could switch between HRTF A and B via buttons A and B. Sitting on the swivel chair, the subject was allowed explorative movements with her head and body and was not instructed to move her head in any particular way [29,46]. The subject was instructed to explore the VR world by switching between the HRTFs and then rate the HRTFs in the respective perceptual quality. The experimenter was in the same semianechoic room as the subject. The subject was given the tasks by the experimenter, for example, "rate the localizability of HRTF A from 1 to 5 and HRTF B from 1 to 5, with 1 being more difficult to localize and 5 easier to localize." The subject then explored the VR world and switched between the HRTFs. As in the study by Hendrix and Barfield [47], no time limit was set, but there was a condition to listen to at least one animation cycle of the drone for each perceptual quality. Once the subject was able to rate the HRTFs, the scene was paused, and the HRTFs were scored. The subject's response was documented by the experimenter. Accordingly, the subject proceeded through the query catalog (Table 1).

Figure 3. Experimental environment in first-person view: winter landscape scene with drone, board for the display of head-related transfer function (HRTF), and touch controllers for switching.



Individualized and Nonindividualized HRTFs

In the second test, the subject rated 3 HRTFs: HRTF X, HRTF Y, and HRTF Z. One of them was their own HRTF again and the other 2 were HRTFs from other people in the database, one of which was very similar to and the other with large deviations from the individualized HRTF of the subject. The left touch controller was used for switching between HRTF X, Y, and Z. The same procedure was followed as in the first test, rating all the attributes (Table 1). The subject left the glasses on during the scoring; all HRTFs were evaluated simultaneously, and the subject could not return to a different attribute. The duration of each test was also documented. In total, 5 perceptual qualities were tested on 5 HRTFs in 1 repetition, ie, 5×5×2 = 50 answers.

In order to find out whether order effects played any role and to obtain a variance within the subject, we performed a repetition with randomized HRTFs. Before the repetition, there was a break of about 10 minutes in which the subject could take off the glasses. The individualized HRTF was rated twice—once in the first and then in the second test. This served as a reference for checking the functionality of the test design. In addition, after the test, subjects were able to comment on further differences apart from rating the HRTFs.

Results

Overview

The evaluation of the attribute localization, externalization, and realism for the first test (individual vs KEMAR) was done with the Wilcoxon signed-rank test and for the second test (individual vs minimal vs maximal) with analysis of variance (ANOVA) and Tukey as ANOVA posthoc analysis. The evaluation of the attributes front-back position was made via the chi-square test with the Fisher test as posthoc and the tone color via the interquartile range.

In order to determine whether different manipulations of stimuli led to different physiological reactions within a group, we applied two-sided *t* tests. Subtests were calculated using *t* tests to investigate possible differences between groups with different expertise and repetition. The groups of different expertise were divided into “low expertise” and “high expertise.” Judgment reliability within the first and the second tests was checked by repetition. There were no significant differences in all perceptual qualities, which meant that with high probability, subjects were able to evaluate all HRTFs reliably (without guessing) with repeated query in spite of randomization. Judgment reliability between the first and second test was assessed by rating the individualized HRTF twice. Evaluation of the individualized HRTF in the first versus the second test showed no significant difference except in externalization, when the *P* value was .04. We attempted to minimize the fatigue effects by the randomized design and a break at halfway through the VR experiment.

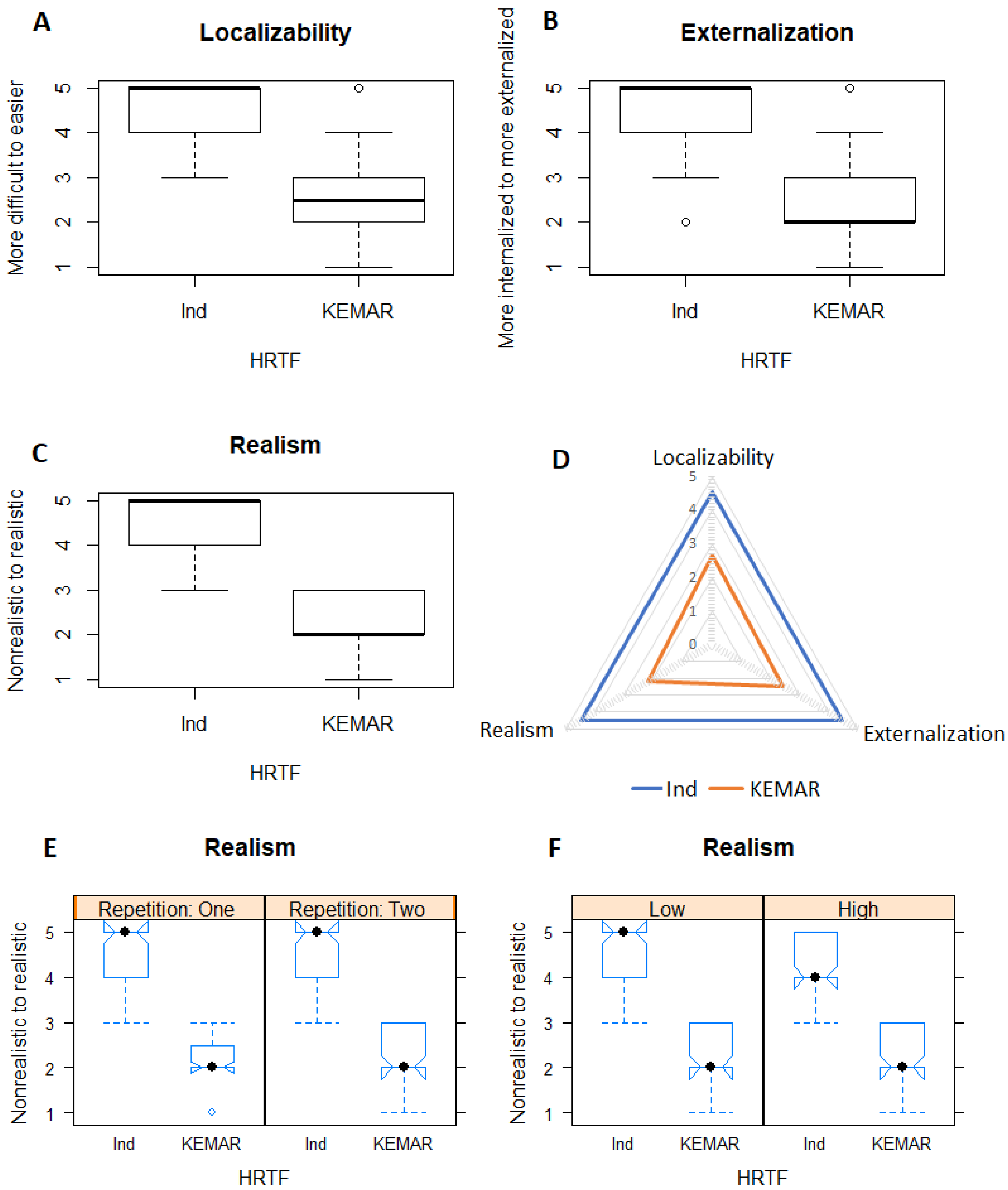
Overall, in both tests, statistical significance was found for all perceptual qualities, except in the front-back position between individual and minimal HRTFs. Plots were calculated and created using the statistics program RStudio (RStudio Inc). In the following section, we offer a detailed statistical analysis of the first and the second tests in terms of the tests themselves and the 5 perceptual qualities (scale of 1-5): localizability (more difficult to easier), front-back position (confused to not confused), externalization (more internalized to more externalized), tone color (darker to brighter), and realism (nonrealistic to realistic).

Individualized and General HRTFs

Localizability

In the assessment of localizability in the first test, the individual HRTF and KEMAR HRTF, which were evaluated with the Wilcoxon signed-rank test, differed significantly (W=5569.5, $P<.001$, Figure 4). The KEMAR HRTF was considered more difficult and the individual HRTF more easily locatable. The subgroup analysis showed great agreement on this result, both between the first and second repetition and in low and high expertise (significance values in Table 2).

Figure 4. Result of the first test (individual vs KEMAR). A: Localizability overall box plot; B: Externalization overall box plot; C: Realism overall box plot; D: Radar chart for localizability, externalization, and realism; E: Realism repetition response behavior; F: Realism expertise response behavior. Ind: individual HRTF; KEMAR: Knowles Electronics Manikin for Acoustic Research; HRTF: head-related transfer function.



XSL•FO
RenderX

Table 2. Significance values of the subgroups for localizability, externalization, and realism in the first test.

Perceptual quality	Individual HRTF[a], *P* value	KEMAR[b] HRTF, *P* value
Localizability repetition	.46	.76
Localizability expertise	.87	.07
Externalization repetition	.26	.71
Externalization expertise	.60	.87
Realism repetition	.84	.22
Realism expertise	.06	.87

[a]HRTF: head-related transfer function.

[b]KEMAR: Knowles Electronics Manikin for Acoustic Research.

Front-Back Position

For the evaluation of the front-back position with its bipolar response scale ("yes" for confused/"no" for not confused), the Pearson's chi-squared test (a frequency test/independence test) with Yates continuity correction was used. The 39 subjects gave 78 responses in 2 rounds. There was no front-back confusion for the individual HRTF (0%). On the other hand, for the KEMAR HRTF, there were 21 front-back confusions reported for 78 responses (27%). The result of the front-back confusions was significantly different ($\chi^2_{1,77}$=22.0, *P*<.001). The KEMAR HRTF was confused by experienced listeners in 14 out of 78 responses (18%) and by inexperienced listeners in 7 out of 78 responses (9%). Due to the possibility of movement by means of head-tracking, basically, no front-back confusion should have occurred [6,37], but the head movements of the subjects were not restricted and therefore could be static as well as dynamic. In practice, movements occur naturally—sometimes more and sometimes less in VR. Nevertheless, with the KEMAR HRTF and a moving stimulus along the sagittal plane, it was still possible that in the VR condition, the visual stimulus was perceived at the front, but the auditory stimulus at the back or vice versa. Wightman and Kistler [37] also detected front-back confusion with uncontrolled sound source movement in their study.

Externalization

The results for the perceived externalization of the test items of the KEMAR and individual HRTF with the Wilcoxon signed-rank test were as follows: the individual HRTF was more significantly externalized than the KEMAR HRTF (W=5741.5, *P*<.001, Figure 4). The subgroups showed no significant differences (Table 2). Although the visual stimulus flew over the head, some subjects reported that the auditory stimulus actually flew through their heads and was thus perceived as being more internalized. This was especially the case when the visual stimulus was just behind the head (not in the field of view). Moreover, when the distance from the auditory stimulus to the visual stimulus was too far, it was rated as 1.

Tone Color

A rating of 1 or 5 meant that the sound of the stimulus was different from what the subject normally perceives (unnaturally brighter or darker). Here, the subjects were supposed to give appropriate answers to the internal reference. The rating of the tone color was difficult for some subjects without a direct reference. Nevertheless, on average, subjects rated the individual HRTF in tone color with 3, which was defined as natural. The individual HRTF was rated as natural in 62 out of 78 responses (79%) and the KEMAR only once (1%). The KEMAR HRTF was mostly rated with 5 in tone color; thus, it was perceived as unnaturally brighter.

Realism

For the final assessment in the first test with the perceptual quality realism, the overall results are shown in Figure 4. The HRTFs were found to be significantly different, with medians of 5 for individual and 2 for KEMAR (W=6030, *P*<.001). There were no significant differences among the subgroups (Table 2).

Individualized and Nonindividualized HRTFs

Localizability

For the evaluation of the localizability in the second test, an ANOVA with Tukey posthoc test was used to compare the 3 HRTFs. In all HRTF scores, we reached the significance levels ($F_{2,76}$=19.131, *P*<.001, Figure 5): individual-maximal (*P*<.001), individual-minimal (*P*=.001), and maximal-minimal (*P*=.049) were calculated using the linear mixed-effects model and the Kenward-Roger method (95% confidence interval). The individual HRTF showed a significantly better behavior than the minimal and maximal HRTFs. In the subgroup, the minimal HRTF was classified as slightly more difficult to localize by the high expertise group (Table 3).

Figure 5. Result of the second test (individual vs minimal vs maximal). A: Localizability overall box plot; B: Externalization overall box plot; C: Realism overall box plot; D: Radar chart for localizability, externalization, and realism; E: Repetition response behavior; F: Expertise response behavior. Ind: individual HRTF; Min: minimal; Max: maximal; HRTF: head-related transfer function.

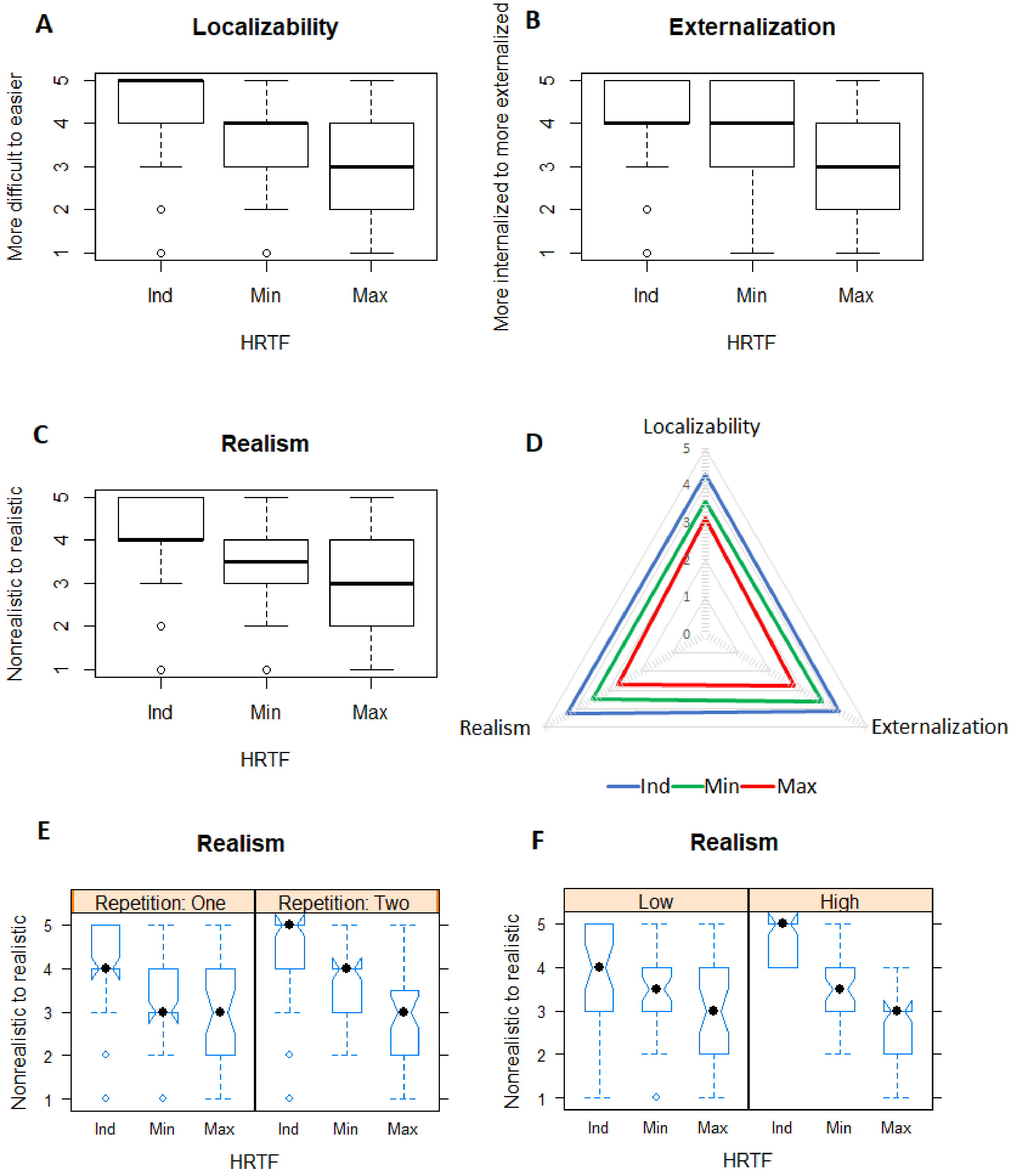


Table 3. Significance values of the subgroups for localizability, externalization, and realism in the second test.

Perceptual quality	Individual HRTF[a], *P* value	Minimal HRTF, *P* value	Maximal HRTF, *P* value
Localizability repetition	.58	.64	.41
Localizability expertise	.06	.03	.32
Externalization repetition	.99	.76	.94
Externalization expertise	.09	.32	.76
Realism repetition	.86	.64	.85
Realism expertise	.04	.78	.29

[a]HRTF: head-related transfer function.

Front-Back Position

The evaluation of the significance level of the yes/no results of the front-back position was calculated with the chi-square test and the Fisher test as chi-square posthoc test ($\chi^2_{2,76}$=27.0, *P*<.001). There was no significant difference between individual and minimal deviant nonindividualized HRTFs (*P*=.06). However, there were significant differences between individual-maximal (*P*<.001) and maximal-minimal (*P*=.005). The application of the Bonferroni method yielded no changes in the significance level: individual-maximal (*P*<.001), individual-minimal (*P*=.18), and maximal-minimal (*P*=.005). For the individual HRTF again (as in the first test), no front-back confusion occurred (0%). However, the minimal HRTF showed front-back confusion in 5 out of 78 responses (6%) and the maximal HRTF in a total of 19 out of 78 responses (24%). The maximal HRTF was therefore rated significantly worse in the front-back position than the other 2 HRTFs. Listeners with high expertise rated the minimal HRTF in 5 out of 78 responses (6%) as reversed in the front-back position and the maximal HRTF in 14 out of 78 responses (18%). Thus, low-expertise listeners could not find any front-back confusion with the minimal HRTF and they found front-back confusion only occasionally (5/78, 6%) with the maximal HRTF. This result of the subgroup was in agreement with the study by Bronkhorst [30] with regard to inexperienced listeners, for whom hardly any front-back confusion errors occurred.

Externalization

For the evaluation of the externalization in the second test, an ANOVA with Tukey posthoc test was used. In all HRTF scores, we achieved significant differences ($F_{2,76}$=22.278, *P*<.001, Figure 5): individual-maximal (*P*<.001), individual-minimal (*P*=.04), and maximal-minimal (*P*<.001) were calculated using the linear mixed effects model and the Kenward-Roger method (95% confidence interval). The maximal HRTF showed a significantly worse behavior than the individual and minimal HRTF. The subgroups showed no significant differences (Table 3).

Tone Color

In order to statistically analyze the evaluation of the tone color—that is how naturally the respective HRTF was perceived—we used descriptive statistics with the interquartile range [48]. Subjects were unanimous in rating their own HRTF according to the small IQR for their individual HRTF. The IQR was the largest at the maximal HRTF (IQR individual HRTF=0, IQR minimal HRTF=1, and IQR maximal HRTF=2.75), which simply indicated that nonindividualized HRTFs were often perceived as unnaturally brighter or darker as well as unpleasant in timbre. Overall, the individual HRTF was rated as natural in 51 out of 78 responses (65%), the minimal HRTF was rated as natural in 23 out of 78 responses (29%), and the maximal HRTF was rated as natural in 15 out of 78 responses (19%); thus, in some cases, nonindividualized HRTFs sometimes were nevertheless rated as natural in tone color.

Realism

All examined HRTFs could be clearly differentiated by rating the attribute realism, of which the overall results are shown in Figure 5. However, the quality of the HRTFs was almost never rated as poor or completely unrealistic. An ANOVA with a Tukey posthoc test showed significant differences in all HRTF scores ($F_{2,76}$=31.88, *P*<.001): individual-maximal (*P*<.001), individual-minimal (*P*<.001), and maximal-minimal (*P*<.001) were calculated using the linear mixed effects model and the Kenward-Roger method (95% confidence interval). The subgroup analysis showed that the individual HRTF was rated more often as being more realistic by the more experienced listeners compared to the low expertise group (Table 3).

After the test, subjects were able to comment on further differences apart from rating the HRTFs. With localizability, subjects reported that they had classified HRTFs as more difficult to localize if the auditory stimulus was not congruent to the visual stimulus and was shifted to the right or left or was diffused. Basically, the group of high expertise found it easier to hear differences between the individual, minimal, and maximal HRTFs than the low expertise group. The first test was classified as being easier for some subjects than the second. Many subjects found the VR scene very realistic, but for some, the auditory stimulus was not a realistic sound to match the drone. NH92 and NH785 had difficulty ignoring the artefacts caused by the lack of interpolation and by the error proneness in the HRTF measurement, but in the end, they rated their own HRTF the highest. All HRTFs were equal because none were interpolated. NH794 perceived individual HRTF as much more realistic and its spectrum much closer to reality. None of the subjects experienced motion sickness during the experiment.

Discussion

Principal Findings

The most important findings of this study are summarized as follows:

1. In VR, there seems to be a connection between auditory spatialization and the descriptive attribute of realism. The perceived realism increases with the approach to listener-specific spatialization.
2. Significant differences in the evaluation of perceptual qualities in VR seem to be mainly caused by listener-specific features. The presentation with individualized HRTFs in VR shows a greater popularity in the subjective rating than with general or nonindividualized HRTFs.
3. The localization model in sagittal planes based on the stationary pure auditory localization error [35] seems to be transferable to the multimodal audiovisual VR. The subjective evaluation reveals the relevance of localization in the dimension of perceived realism. Even HRTFs with a localization error that only deviates minimally in the static auditory are evaluated as less realistic in a direct comparison with their own HRTFs in a complex scene in a multimodal representation.

Comparison With Hypotheses

Contrary to our expectations, the use of the tracking system and visual stimuli did not significantly reduce the number of front-back confusions for the KEMAR and maximal HRTF. Furthermore, the results prove the following concerning our hypotheses.

The first hypothesis of our study that general HRTFs lead to limitations of 3D audio perception in VR was confirmed. The first test (individual vs KEMAR HRTF) showed that subjects with a general artificial head HRTF had more difficulties locating moving sound sources in VR. They were confused in the front-back position, found it to be more internalized, and rated the tone color as unnatural and unrealistic. However, the test was performed only with a general HRTF—the KEMAR HRTF from the most widely used MIT KEMAR database [36]—and is thus valid only for this HRTF. In order to make a global statement, several general HRTFs, for instance, that of Neumann KU100, should be included in the investigation. Moreover, the comparison between individual and KEMAR HRTF can be criticized, as the resolution of the individual HRTF was better, with 1550 positions and 256 samples at 48 kHz, compared to the KEMAR HRTF with 710 positions and 512 samples at 44.1 kHz, although the sampling rate had been adjusted. The resolution could therefore be another factor, but it is obvious that the KEMAR HRTF, which has been used almost exclusively in games with spatialization, leads to limitations in 3D audio perception. Although we considered a downgrade of the individual HRTF, no consensus with the same position measurement points could be found.

The second hypothesis on the transferability of the localization model for stationary localization errors for nonindividualized HRTFs in more complex environments such as VR was unequivocally confirmed with the second test of this study. Admittedly, the HRTF selection could have been even more specific to determine the correlation between the localization error and perceived realism in truly complex environments. In the future, the deviations of the stationary localization errors could be represented by means of the model in percentage: for instance, individualized HRTFs would have a localization error deviation of 0% and nonindividualized HRTFs with increasing inaccuracy at an increasing value of up to 100% (=error at random localization). Thus, more than 2 nonindividualized HRTFs could be selected at fixed percentages and a finer resolution of the degree of realism compared to the localization error would be possible. Moreover, other databases could have been included, but we wanted to maintain the comparability of the measurements. Finally, only 2 nonindividualized HRTFs were selected for the second test because the number of HRTFs in the database was too small to make general statements.

Comparison With Prior Work

Studies such as those of Begault et al, Hendrix and Barfield, and Larsson et al [29,47,49] have already tried to examine the relationship between realism and improving spatialization with HRTF rendering, but they found no significant differences. This is probably due to the lack of understanding or the unclear definition of what is meant by the assessment of realism. The explanation by Hendrix and Barfield [47] for their findings was that the subjects might have interpreted the realism in terms of the visual realism "scene realism" and not the overall quality of the performance. Additionally, Begault et al [29] argued that no differences in realism were found, because subjects probably had no common understanding of what the perceived realism implied. Furthermore, Larsson et al [49] did not define the queried realism in advance and suggested that the subjects had made the auditive realism more dependent on well-designed source content (eg, a bus really sounds like a bus) instead of on one accurate 3D performance (that the bus is properly externalized and located). In our study, the concept of realism was defined in advance according to the study by Simon et al [39]. Thus, a common understanding of the queried realism was guaranteed for all listeners.

By examining the differences in the perceptions between individualized and general or nonindividualized HRTFs in VR, reproduction systems are to be examined in order to generate virtual and augmented realities as realistically as possible. Unlike in the study by Berger et al [50], in this study, limitations in 3D audio perception with general HRTFs in VR arose. It is questionable whether this claim in the title of the study carried out by Microsoft Research with the MIT KEMAR HRTF dataset alone is justified without comparison to individualized HRTFs. Further, it should be noted that the Berger study did not evaluate elevation but only azimuth.

Another point is the learnability or the adaptability of foreign HRTFs. Studies [2] have shown that HRTFs can be learned and adapted through training in a short amount of time, sometimes even within minutes [51], but this has only been evaluated by localization performance. Whether an adaptation in the evaluation of externalization, tone color, or realism in VR is possible remains a question. Presumably, learning new HRTFs comprehensively is a longer process and more akin to learning a foreign language.

Current research especially for serious games in VR and mental health often mention that “The literature suggests that immersion is largely influenced by both visual and audio qualities” [52], but audio is rarely a topic in such studies. “VR excels in its advantage of being able to draw on both audio and interactive visual stimuli, making the fearful stimuli appear as real as possible [53],” while by using personalized HRTFs, the stimuli would be even more present, immersive, and realistic.

Limitations

In a realistic virtual scene, the reverberation should not be neglected. This could be set in the reference setup via the plugin (Plugin Spatializer Reverb in the Audio Mixer, [40]) as well as via the Unity Engine (Audio Reverb Filter, Reverb Preset). However, the reverberation leads to another variable with many parameters. In our study, all effects were therefore examined in free field. This may sound unrealistic but this corresponds to a real situation in a room that is acoustically dry. For further experiments, the reverberation can be integrated as an additional variable building on this work.

Conclusions

Both hypotheses have been accepted: first, general HRTFs lead to limitations of 3D audio perception in VR and second, the localization model for stationary localization errors is transferable to nonindividualized HRTFs in more complex environments such as VR. The results of the first test (individual vs KEMAR HRTF) and of the second test (individual vs minimal vs maximal) show that sounds filtered by individualized HRTFs are considered easier to localize, easier to externalize, more natural in timbre, and thus more realistic compared to sounds filtered by nonindividualized HRTFs. In conclusion, the most realistic simulation of sound sources in virtual environments can be achieved by using individualized HRTFs, which leads to an improvement in terms of the following perceptual qualities: localizability, front-back position, externalization, tone color, and realism. Therefore, future VR studies, especially in serious games, should take an auditory spatialization with individual HRTFs in their experiments into account.

To answer the question “Binaural Technique: Do We Need Individual Recordings?” by Møller et al [1] in the field of VR, this study provides empirical evidence. The answer is in the affirmative. Listener-specific filtering in headphone reproduction helps achieve a truly realistic 3D audio perception in VR. In order to see the topic of the necessity of a higher realism content in VR by means of individual HRTFs, less from theoretical basic research and more from the side of practical realization, the following example provides a nice vivid comparison: HRTFs are like a suit. It fits you perfectly when it is tailor-made.

Acknowledgments

We thank Piotr Majdak, Bernhard Laback, Robert Baumgartner, and Sridhar Srinivasan for their numerous competent suggestions, Michael Mihocic for his help in running the HRTF measurements and Christian Kasess, Nicola Klingler, Hannah Leykum, and Marik Roos for their fruitful discussions on statistics. Open access funding was provided by the University of Vienna.

Conflicts of Interest

None declared.

References

1. Møller H, Sørensen MF, Jensen CB, Hammershøi D. Binaural Technique: Do We Need Individual Recordings? J Audio Eng Soc 1996;44(6):451-469 [FREE Full text]
2. Mendonça C. A review on auditory space adaptations to altered head-related cues. Front Neurosci 2014;8:219 [FREE Full text] [doi: 10.3389/fnins.2014.00219] [Medline: 25120422]
3. Kolarik AJ, Moore BCJ, Zahorik P, Cirstea S, Pardhan S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Atten Percept Psychophys 2016 Feb;78(2):373-395 [FREE Full text] [doi: 10.3758/s13414-015-1015-1] [Medline: 26590050]
4. Hartmann WM. Localization of sound in rooms. J Acoust Soc Am 1983;74(5):1380-1391. [doi: 10.1121/1.390163] [Medline: 6643850]
5. Vorländer M. Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Berlin Heidelberg: Springer; 2008.
6. Carlile S, Leung J. The Perception of Auditory Motion. Trends Hear 2016;20 [FREE Full text] [doi: 10.1177/2331216516644254] [Medline: 27094029]
7. Bertelson P, Radeau M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept Psychophys 1981;29(6):578-584. [doi: 10.3758/bf03207374] [Medline: 7279586]
8. Carlile S, Fox A, Orchard-Mills E, Leung J, Alais D. Six Degrees of Auditory Spatial Separation. J Assoc Res Otolaryngol 2016;17(3):209-221 [FREE Full text] [doi: 10.1007/s10162-016-0560-1] [Medline: 27033087]
9. Stein BE, editor. The New Handbook of Multisensory Processing. Cambridge, MA: MIT Press; 2012.
10. Thorne JD, Debener S. Look now and hear what's coming: on the functional role of cross-modal phase reset. Hear Res 2014;307:144-152. [doi: 10.1016/j.heares.2013.07.002] [Medline: 23856236]

11. Middlebrooks JC. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. J Acoust Soc Am 1999;106(3):1493-1510. [doi: 10.1121/1.427147] [Medline: 10489706]
12. Wenzel EM, Arruda M, Kistler DJ, Wightman FL. Localization using nonindividualized head-related transfer functions. J Acoust Soc Am 1993;94(1):111-123. [doi: 10.1121/1.407089] [Medline: 8354753]
13. Snoswell AJ, Snoswell CL. Immersive Virtual Reality in Health Care: Systematic Review of Technology and Disease States. JMIR Biomed Eng 2019 Sep 26;4(1):e15025. [doi: 10.2196/15025]
14. Ijaz K, Ahmadpour N, Naismith SL, Calvo RA. An Immersive Virtual Reality Platform for Assessing Spatial Navigation Memory in Predementia Screening: Feasibility and Usability Study. JMIR Ment Health 2019;6(9):e13887 [FREE Full text] [doi: 10.2196/13887] [Medline: 31482851]
15. Otkhmezuri B, Boffo M, Siriaraya P, Matsangidou M, Wiers RW, Mackintosh B, et al. Believing Is Seeing: A Proof-of-Concept Semiexperimental Study on Using Mobile Virtual Reality to Boost the Effects of Interpretation Bias Modification for Anxiety. JMIR Ment Health 2019;6(2):e11517 [FREE Full text] [doi: 10.2196/11517] [Medline: 30789353]
16. Ose SO, Færevik H, Kaasbøll J, Lindgren M, Thaulow K, Antonsen S, et al. Exploring the Potential for Use of Virtual Reality Technology in the Treatment of Severe Mental Illness Among Adults in Mid-Norway: Collaborative Research Between Clinicians and Researchers. JMIR Form Res 2019;3(2):e13633 [FREE Full text] [doi: 10.2196/13633] [Medline: 31199315]
17. Akeroyd MA. An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. Trends Hear 2014;18 [FREE Full text] [doi: 10.1177/2331216514560442] [Medline: 25492094]
18. Blauert J. Sound Localization in the Median Plane. Acustica 1969;22:205-213 [FREE Full text]
19. Middlebrooks JC. Sound localization. In: Handbook of Clinical Neurology. Amsterdam: Elsevier; 2015:99-116.
20. Middlebrooks JC, Green DM. Sound localization by human listeners. Annu Rev Psychol 1991;42:135-159. [doi: 10.1146/annurev.ps.42.020191.001031] [Medline: 2018391]
21. Mills AW. Auditory Localization. In: Tobias JV, editor. Foundations of modern auditory theory. New York: Academic Press; 1972:303-348.
22. Lord Rayleigh WS. On our perception of sound direction. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 1907;13(74):214-232. [doi: 10.1080/14786440709463595]
23. Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. J Acoust Soc Am 2002;111(5):2219-2236. [doi: 10.1121/1.1471898] [Medline: 12051442]
24. Blauert J. Spatial hearing: The Psychophysics of Human Sound Localization. 2nd edition. Cambridge, MA: MIT-Press; 1997.
25. Møller H, Sørensen MF, Hammershøi D, Jensen CB. Head-related transfer functions of human subjects. J Audio Eng Soc 1995;43(5):300-321 [FREE Full text]
26. Xie B. Head-related transfer function and virtual auditory display. Plantation, FL: J. Ross Publishing; 2013.
27. Morimoto M, Ando Y. On the simulation of sound localization. J Acoust Soc Japan 1980;1(3):167-174. [doi: 10.1250/ast.1.167]
28. Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: Psychophysical validation. J Acoust Soc Am 1989;85(2):868-878. [doi: 10.1121/1.397558] [Medline: 2926001]
29. Begault DR, Wenzel EM, Anderson MR. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J Audio Eng Soc 2001 Oct;49(10):904-916. [Medline: 11885605]
30. Bronkhorst AW. Localization of real and virtual sound sources. The Journal of the Acoustical Society of America 1995 Nov;98(5):2542-2553. [doi: 10.1121/1.413219]
31. Durlach NI, Rigopulos A, Pang XD, Woods WS, Kulkarni A, Colburn HS, et al. On the Externalization of Auditory Images. Presence: Teleoperators & Virtual Environments 1992;1(2):251-257. [doi: 10.1162/pres.1992.1.2.251]
32. Hartmann WM, Wittenberg A. On the externalization of sound images. J Acoust Soc Am 1996;99(6):3678-3688. [doi: 10.1121/1.414965] [Medline: 8655799]
33. Kim S, Choi W. On the externalization of virtual sound images in headphone reproduction: a Wiener filter approach. J Acoust Soc Am 2005;117(6):3657-3665. [doi: 10.1121/1.1921548] [Medline: 16018469]
34. Romigh GD, Simpson BD. Do you hear where I hear?: isolating the individualized sound localization cues. Front Neurosci 2014;8. [doi: 10.3389/fnins.2014.00370]
35. Baumgartner R, Majdak P, Laback B. Modeling sound-source localization in sagittal planes for human listeners. J Acoust Soc Am 2014;136(2):791-802 [FREE Full text] [doi: 10.1121/1.4887447] [Medline: 25096113]
36. Gardner WG, Martin KD. HRTF measurements of a KEMAR. J Acoust Soc Am 1995;97(6):3907-3908. [doi: 10.1121/1.412407]
37. Wightman FL, Kistler DJ. Resolution of front-back ambiguity in spatial hearing by listener and source movement. J Acoust Soc Am 1999;105(5):2841-2853. [doi: 10.1121/1.426899] [Medline: 10335634]
38. Lindau A, Erbes V, Lepa S, Maempel HJ, Brinkman F, Weinzierl S. A Spatial Audio Quality Inventory (SAQI). Acta Acustica united with Acustica 2014;100(5):984-994. [doi: 10.3813/aaa.918778]

39. Simon LSR, Zacharov N, Katz BFG. Perceptual attributes for the comparison of head-related transfer functions. J Acoust Soc Am 2016;140(5):3623-3632. [doi: 10.1121/1.4966115] [Medline: 27908072]
40. Jenny C, Majdak P, Reuter C. SOFA Native Spatializer Plugin for Unity - Exchangeable HRTFs in Virtual Reality. Presented at: Proceedings of the 144th Convention of the Audio Eng Soc, Convention e-brief 406; 2018; Milano, Italy URL: https://www.aes.org/e-lib/browse.cfm?elib=19519
41. Kopiez R, Wolf A, Platz F, Mons J. Replacing the Orchestra? - The Discernibility of Sample Library and Live Orchestra Sounds. PLoS One 2016;11(7):e0158324 [FREE Full text] [doi: 10.1371/journal.pone.0158324] [Medline: 27382932]
42. Majdak P, Goupell MJ, Laback B. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. Atten Percept Psychophys 2010;72(2):454-469 [FREE Full text] [doi: 10.3758/APP.72.2.454] [Medline: 20139459]
43. Majdak P, Balazs P, Laback B. Multiple exponential sweep method for fast measurement of head-related transfer functions. J Audio Eng Soc 2007;55(7/8):623-637 [FREE Full text]
44. Søndergaard P, Majdak P. The Auditory Modeling Toolbox. In: Blauert J, editor. The Technology of Binaural Listening. Berlin Heidelberg New York: Springer; 2013:33-56.
45. Majdak P, Carpentier T, Nicol R, Roginska A, Suzuki Y, Watanabe K, et al. Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions. Presented at: Proceedings of the 134th Convention of the Audio Eng Soc, Convention Paper 8880; 2013; Roma, Italy URL: https://www.aes.org/e-lib/browse.cfm?elib=16781
46. Romigh GD, Brungart DS, Simpson BD. Free-Field Localization Performance With a Head-Tracked Virtual Auditory Display. IEEE J Sel Top Signal Process 2015;9(5):943-954. [doi: 10.1109/jstsp.2015.2421874]
47. Hendrix C, Barfield W. The Sense of Presence within Auditory Virtual Environments. Presence: Teleoperators & Virtual Environments 1996;5(3):290-301. [doi: 10.1162/pres.1996.5.3.290]
48. Lindau A, Brinkmann F, Weinzierl S. Sensory Profiling of Individual and Non-individual Dynamic Binaural Synthesis Using the Spatial Audio Quality Inventory. Presented at: Proceedings of the EAA Forum Acusticum; 2014; Kraków URL: https://tinyurl.com/y58t933n
49. Larsson P, Västfjäll D, Kleiner M. Effects of auditory information consistency and room acoustic cues on presence in virtual environments. Acoust Sci & Tech 2008;29(2):191-194. [doi: 10.1250/ast.29.191]
50. Berger CC, Gonzalez-Franco M, Tajadura-Jiménez A, Florencio D, Zhang Z. Generic HRTFs May be Good Enough in Virtual Reality. Improving Source Localization through Cross-Modal Plasticity. Front Neurosci 2018;12:21 [FREE Full text] [doi: 10.3389/fnins.2018.00021] [Medline: 29456486]
51. Parseihian G, Katz BFG. Rapid head-related transfer function adaptation using a virtual auditory environment. J Acoust Soc Am 2012;131(4):2948-2957. [doi: 10.1121/1.3687448] [Medline: 22501072]
52. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual Reality Clinical Research: Promises and Challenges. JMIR Serious Games 2018;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
53. Jerdan SW, Grindle M, van Woerden HC, Kamel Boulos MN. Head-Mounted Virtual Reality and Mental Health: Critical Review of Current Research. JMIR Serious Games 2018;6(3):e14 [FREE Full text] [doi: 10.2196/games.9226] [Medline: 29980500]

Abbreviations

ANOVA: analysis of variance
ARI: Acoustic Research Institute
HRTF: head-related transfer function
KEMAR: Knowles Electronics Manikin for Acoustic Research
MIT: Massachusetts Institute of Technology
NH: normal hearing listener
PE: polar error
QE: quadrant error
SOFA: Spatially Oriented Format for Acoustics
VR: virtual reality

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by M Eckert, A Zaidi; comments to author 20.04.20; revised version received 07.05.20; accepted 26.07.20; published 08.09.20.

Please cite as:
Jenny C, Reuter C
Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study With Perceptual Attributes in Sagittal Plane Sound Localization
JMIR Serious Games 2020;8(3):e17576
URL: http://games.jmir.org/2020/3/e17576/
doi:10.2196/17576
PMID:32897232

Original Paper

An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study

Dieter Lerner[1], MA; Stefan Mohr[2], MD; Jonas Schild[3], PhD; Martin Göring[2], MD; Thomas Luiz[1], MD

[1]Fraunhofer Institute for Experimental Software Engineering, Research Program Digital Healthcare, Kaiserslautern, Germany

[2]University Hospital Heidelberg, Clinic of Anesthesiology, Heidelberg, Germany

[3]Hannover University of Applied Sciences and Arts, Interactive Reality Experiences, Hannover, Germany

Corresponding Author:
Thomas Luiz, MD
Fraunhofer Institute for Experimental Software Engineering
Research Program Digital Healthcare
Fraunhofer-Platz 1
Kaiserslautern, 67663
Germany
Phone: 49 631 6800 2148
Email: Thomas.Luiz@iese.fraunhofer.de

Abstract

Background: Virtual reality (VR) is increasingly used as simulation technology in emergency medicine education and training, in particular for training nontechnical skills. Experimental studies comparing teaching and learning in VR with traditional training media often demonstrate the equivalence or even superiority regarding particular variables of learning or training effectiveness.

Objective: In the EPICSAVE (Enhanced Paramedic Vocational Training with Serious Games and Virtual Environments) project, a highly immersive room-scaled multi-user 3-dimensional VR simulation environment was developed. In this feasibility study, we wanted to gain initial insights into the training effectiveness and media use factors influencing learning and training in VR.

Methods: The virtual emergency scenario was anaphylaxis grade III with shock, swelling of the upper and lower respiratory tract, as well as skin symptoms in a 5-year-old girl (virtual patient) visiting an indoor family amusement park with her grandfather (virtual agent). A cross-sectional, one-group pretest and posttest design was used to evaluate the training effectiveness and quality of the training execution. The sample included 18 active emergency physicians.

Results: The 18 participants rated the VR simulation training positive in terms of training effectiveness and quality of the training execution. A strong, significant correlation (r=.53, *P*=.01) between experiencing presence and assessing training effectiveness was observed. Perceived limitations in usability and a relatively high extraneous cognitive load reduced this positive effect.

Conclusions: The training within the virtual simulation environment was rated as an effective educational approach. Specific media use factors appear to modulate training effectiveness (ie, improvement through "experience of presence" or reduction through perceived limitations in usability). These factors should be specific targets in the further development of this VR simulation training.

(JMIR Serious Games 2020;8(3):e18822) doi:10.2196/18822

KEYWORDS

virtual reality; educational virtual realities; virtual patient simulation; virtual emergency scenario; simulation training; head-mounted display; immersive media; training effectiveness; presence experience; anaphylaxis; emergency medicine

Introduction

Simulation Training in Emergency Medicine

Working in emergency medicine is characterized by constraints that constitute a high-risk constellation: need for situational assessment and decision making as well as initiation of appropriate emergency measures under time pressure, often under adverse external conditions and, at the same time, with little or no fault tolerance [1]. In the continuing education and training of emergency physicians, simulation therefore plays an ever-greater role than in other contexts. Usually, complex full-scale simulators with audio-video systems and extensive emergency equipment are required for the training of emergency physicians [2]. Initial studies report very high appreciation by the participants of the practical relevance, preparation for real-life emergency missions, and learning effect of this kind of training within the German Emergency Medical Services System [3].

Virtual Reality Simulation

Virtual reality (VR) is increasingly used as simulation technology in emergency medicine, in particular for training nontechnical skills [4]. Experimental studies conducted with health care professionals, where training in VR is compared with traditional training media, often demonstrate the equivalence or even superiority of VR regarding learning or training effectiveness [5]. Highly immersive VR (ie, VR supported by VR glasses) enable spatial positioning of the users in a virtual 3-dimensional (3D) simulation environment and the feeling of being present in the VR. A high level of experiencing presence correlates positively with variables for learning or training effectiveness [6]. These immersive VR situations are capable of presenting visual information in a more realistic, 3D manner, either by providing a truly immersive experience or by presenting standardized and repeatable environments and patients that are otherwise not accessible, safe, "tangible," or ethical. In the near future, immersive VR could be integrated regularly into existing simulation training as a complementary arrangement [7]. However, media use factors can also reduce perceived training effectiveness (eg, perceived usability) [8].

Objectives

The goal of this feasibility study was to gain initial insights into the training effectiveness of a highly immersive multi-user VR simulation environment. Our intention was not to advocate VR simulation training as a replacement for traditional simulation training, but rather as a complement for training nontechnical skills (eg, clinical and procedural reasoning). The issues investigated in this study were: (1) How do emergency physicians assess the overall training effectiveness? and (2) How do they assess further media-specific variables influencing training in VR, such as the experience of presence, usability, cognitive load (CL), VR sickness, and intrinsic motivation?

Methods

Project Background

In the EPICSAVE (Enhanced Paramedic Vocational Training with Serious Games and Virtual Environments) project, a highly immersive room-scaled, multi-user, 3-dimensional simulation environment was developed for the team training of emergency medical services personnel as a complement to traditional simulation training. The higher-level learning goals were to improve clinical reasoning, procedural reasoning, and cooperation in the team [9]. Within the VR training environment, acquisition of practical skills (ie, insertion of intravenous cannula) is neither intended nor technically possible. Following early evaluation studies with future paramedics [10], the system was used with emergency physicians to validate the transfer to higher levels of expertise.

EPICSAVE VR Simulation Environment

For team training with 2 participants, the hardware equipment consisted of 2 sets of VR glasses with integrated headphones (head-mounted displays; HTC Vive Pro), 4 input devices (HTC Controller) for the selection of menu items or for activating and moving objects, and 2 PCs for controlling the simulation. The interactions between the training participants were transmitted to the VR in real time by means of a motion-tracking system and were executed by their virtual representatives (avatars). This enabled collaborative activities in the VR (eg, handing over virtual equipment). The trainer monitored the actions of the participants in the VR via the PC monitors (Figure 1).

Figure 1. Two training participants equipped with virtual reality glasses and input devices.

An integrated virtual patient (VP) represented a multitude of vital parameters and symptom characteristics, which went beyond what can currently be represented by commercially available high-fidelity simulators (Textbox 1). Intravenous infusions or oxygen supplies emptied in real time according to the selected settings. The control of vital parameters and symptom characteristics could either be automated by the trainer using so-called presets or exercised "manually" via an easy-to-use editor. At the time the evaluation was performed, the VP was not yet equipped with language recognition and output, which is why answers from the VP were provided by the trainer off-screen.

Textbox 1. Symptom characteristics and measures of the virtual patient.

Controllable symptom characteristics of the virtual patient

- Psychomotor condition
- Vital parameters (level of consciousness; respiratory rate, depth, and pattern; pulse rate and force; capillary refill time)
- Pathological breathing sounds
- Swelling of skin and mucous membranes
- Exanthema
- Urticaria
- Cyanosis
- Perspiration

Diagnostic and therapeutic measures

- Change of posture
- Undressing
- Monitor-based measurements (electrocardiogram, noninvasive automatic blood pressure, pulse oximetry, temperature, blood glucose)
- Clinical examination (inspection, palpation, auscultation)
- Administration of oxygen
- Application of infusions and medication (eg, inhalative, intravenous)
- Defibrillation

The VR system recorded the training sessions. Important diagnostic steps or therapeutic interventions were automatically recognized, displayed to the trainer in a window on their PC monitor, and presented in chronological and systematic order for the debriefing.

Sample and Study Design

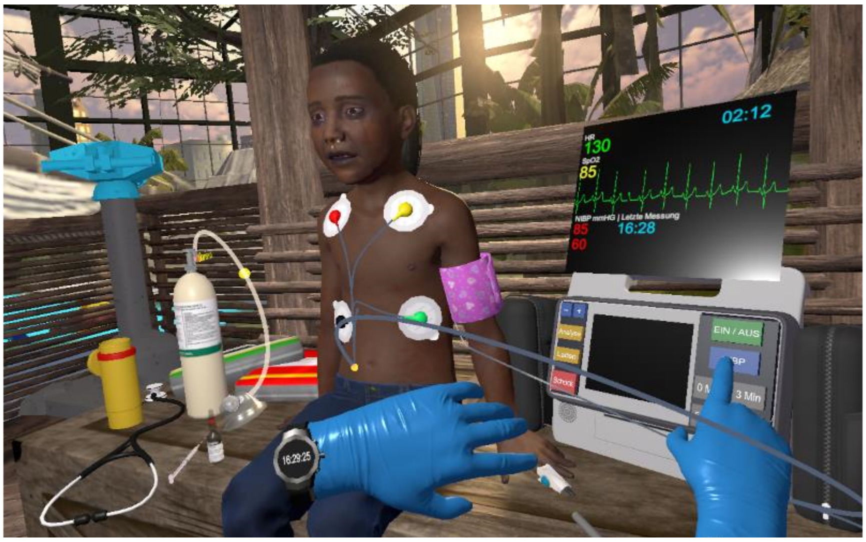
Sample

The 18 participating emergency physicians (14 men, 4 women) for the field study were recruited from the Clinic of Anesthesiology of the Heidelberg University Hospital following a hospital-internal written and verbal call for participation. The mean age of the participants was 36.6 years (SD 7.1 years). The mean years of practical experience was 6.4 years (SD 7.1 years). None of the participants had experience with a highly immersive VR simulation environment prior to training. They were informed about the specific emergency scenario within the training session during the briefing phase. Exclusion and discontinuation criteria for participation in the one-group pretest and posttest design study were VR sickness symptoms such as discomfort, headache, or nausea.

Emergency Scenario

The virtual emergency scenario was anaphylaxis grade III [11] with shock, swelling of the upper and lower respiratory tract, and skin symptoms in a 5-year-old girl (VP) visiting an indoor family amusement park (Figure 2). This clinical scenario was chosen because studies have shown that the potentially life-saving epinephrine therapy prioritized in the guidelines is very often carried out too late or incorrectly in practice [12,13]. Furthermore, pediatric emergencies are associated with very high levels of emotional stress for emergency medical services personnel [14].

Figure 2. The virtual patient in the virtual environment.



Training Session

The VR training was based on the basic principle of simulation-based training [15]: briefing and VR familiarization phase (approximately 20 minutes), training (approximately 20 minutes), and debriefing (approximately 15 minutes). The team consisted of 2 physicians, while the concrete composition of the team was random. The training session was led by a VR system administrator and an experienced Advanced Life Support and Crew Resource Management Training Instructor. The central task during the training was to make a clinical decision regarding working diagnosis and therapy using the widespread Airway, Breathing, Circulation, Disability, Exposure (ABCDE) method and SAMPLER (Signs/symptoms, Allergies, Medication, Past medical history, Last meal, Events prior to incident, Risk factors) scheme while obeying the principles of effective teamwork and team communication. As part of the debriefing phase, immediate feedback for each team was given specifically on the working diagnosis, procedures for communication and teamwork, and epinephrine dose administered.

Data Collection

Measurement of Training Effectiveness

Perceived training effectiveness was measured with the Training Evaluation Inventory (TEI), with the variable values ranging from 1 to 5 [16]. In addition, pretest and posttest scores of declarative and procedural knowledge were collected. The knowledge test (see Multimedia Appendix 1) consisted of 20 multiple choice questions on declarative and procedural knowledge regarding the current guidelines for prehospital diagnosis and therapy of anaphylaxis [11], with a value range of 0-20 according to the multiple-choice typology of Krebs [17]. The validity of the content was checked by 4 emergency physicians with clinical and teaching experience. A higher score corresponds to a higher level of knowledge [17].

Measurement of Clinical Reasoning Skills

We integrated a checklist into the VR training system as an internal embedded assessment system with the following performance data: completeness, order, and time of the ABCDE method; time until the diagnostic decision was made; time until the decision to administer adrenaline; and doses of adrenaline. Not all performance parameters could be recorded automatically due to technical problems with the VR tracking system. Therefore, some data had to be entered manually during the training. Unfortunately, we were unable to ensure that this manually entered data was precise or recorded in a timely manner. This poor reliability prevented us from using this assessment data.

Measurement of Additional Variables Regarding Media Use

The Igroup Presence Questionnaire (IPQ) was used to measure the experience of presence, with the variable values ranging from 0 to 6 [18]. The System Usability Scale (SUS) was used for measuring perceived usability, with the variable values ranging from 0 to 100 [19]. Perceived CL was measured with the scale by Klepsch and colleagues [20], with the variable values ranging from 0 to 7. The scale measures different aspects of cognitive load, namely intrinsic (ICL), extraneous (ECL), and germane (GCL) cognitive load. The mean values for ICL and GCL should normally be higher than those for ECL. The ECL captures the extent to which trainees engage with learning-irrelevant information during training [20]. VR sickness was assessed with a single modified item from the Simulator Sickness Questionnaire [21]. The item for assessing pretest and posttest VR sickness consisted of the question: Do you feel general discomfort (eg, symptoms such as tiredness, headache, nausea, dizziness, or difficulty with visual acuity)? The possible values were: 0 (not at all), 1 (little), 2 (perceptible), and 3 (strong). Pretest and posttest intrinsic motivation was captured with the Situational Motivation Scale (SIMS), with variable values ranging from 4 to 28, where a higher score corresponds to a higher level of intrinsic motivation [22].

Once the participants had been briefed and their consent had been obtained, the pretest was performed. Immediately after the training, the posttest was performed. For the primary data set of pretest data, see Multimedia Appendix 2; for the primary data set of posttest data, see Multimedia Appendix 3.

Data Analysis

The data were collected in pseudonymized form and analyzed and stored in anonymized form. Bivariate and linear correlation analyses were the primarily forms of analyses. Pretest and posttest changes were assessed by means of the corresponding significance tests in accordance with the methodological literature. Regarding statistical and analysis software, SPSS (version 19) and G*Power (version 3.1.9.2) were used. The significance level for a type I error (α) was set at .05.

Ethics Approval and Consent to Participate

The study protocol was approved by the Ethics Commission of the Medical Faculty of the University of Heidelberg, Germany (no. S-646/2018).

The participants were recruited from the Clinic of Anesthesiology of the Heidelberg University Hospital following a hospital-internal written call for participation. They were informed in writing and orally, and their written consent was obtained for the study and for publication of the study results.

Results

Training Effectiveness

Participant-assessed training effectiveness (TEI) resulted in a mean score of 4.38 points (SD 0.41 points). The knowledge test performed right after the conclusion of the training revealed a small, nonsignificant increase in knowledge compared to the baseline value (Table 1).

Table 1. Comparison of pretest and posttest knowledge test results (total score range, 0-20).

Sample	Pretest, mean (SD)	Posttest, mean (SD)	Statistical comparison			
			t test	*P* value[a]	*d*	beta
n=18	14.72 (1.74)	15.50 (1.77)	t_{17}=1.79	.09	0.42	.50

[a]Two-sided *t* test.

Variables of Media Use

Concerning the experience of presence (IPQ), the overall mean score was 3.79 points (SD 0.56 points). In terms of the relationship between the experience of presence (IPQ) and the TEI subscales, a significant correlation was observed mainly for the training effectiveness dimensions "Subjective enjoyment," "Perceived usefulness," and "Attitude towards training" (Table 2).

The bivariate correlation analysis revealed a large, significant correlation between experience of presence (IPQ) and perceived overall training effectiveness (TEI): *r*=.53, *P*=.01. In the linear regression analysis, a comparably large effect could also be seen: R^2=.240, $F_{1,16}$=6.368, *P*=.02, *f*=.56.

Participant-assessed usability (SUS) resulted in a mean score of 65.56 points (SD 14.54 points). In terms of the relationship between experience of presence (IPQ) and usability (SUS), a significant correlation (*r*=.44, *P*=.03) was observed.

The assessment of CL yielded the following results for the respective subdimensions: ICL (mean 3.75, SD 1.31), GCL (mean 3.74, SD 1.36), and ECL (mean 4.39, SD 1.63).

The pretest (mean 0.44, SD 0.51) and posttest (mean 0.44, SD 0.86) scores for VR sickness (modified item from the Simulator Sickness Questionnaire) were identical. Only one person scored his or her sickness as a 2 (perceptible), and one person scored his or her sickness as a 3 (strong) after the training, but neither of these participants had to stop the training because of VR sickness.

Compared to the baseline value, intrinsic motivation (SIMS) increased significantly after the training (Table 3).

The bivariate correlation analysis revealed a strong, significant correlation between experience of presence (IPQ) and situational, intrinsic motivation (SIMS) after the VR simulation training: *r*=.62, *P*=.003. In the linear regression analysis, a comparably high effect was visible: R^2=.344, $F_{1,16}$=9.931, *P*=.01, *f*=.72.

Table 2. Correlations between the Igroup Presence Questionnaire (IPQ; total score range, 0-6) and Training Evaluation Inventory (TEI) subscales (total score range, 1-5).

TEI subscales	Mean	SD	*r*	*P* value
a. Subjective enjoyment	4.67	0.40	.43	.04
b. Perceived usefulness	4.25	0.69	.52	.01
c. Perceived difficulty	4.71	0.37	.11	.34
d. Subjective knowledge gain	3.76	0.75	.39	.06
e. Attitude towards training	4.43	0.47	.54	.01

Table 3. Pretest and posttest results of the Situational Motivation Scale (SIMS; total score range, 4-28).

Sample	Pretest, mean (SD)	Posttest, mean (SD)	Statistical comparison			
			t test	*P* value[a]	*d*	beta
n=18	20.18 (4.62)	22.39 (3.42)	t_{17}=2.60	.02	0.61	.52

[a]Two-sided *t* test.

Discussion

Principal Findings

The highly immersive virtual simulation environment enables dynamically changeable, realistic, and 3D visualization of different clinical environments and VPs from different perspectives. To get initial insights into the potential for using this VR system for simulation training, in this feasibility study, we evaluated the perceived training effectiveness and assessed further media use variables. The results regarding TEI were above average, meaning that the new VR training was not perceived as inferior.

We observed a strong, significant correlation between experience of presence and perceived training effectiveness. The values measured by the IPQ showed a rather high degree of presence experience by the training participants. A high experience of presence is highlighted in the literature as an indicator of highly interactive media, an indicator of systematic cognitive engagement with the contents of the virtual environment, and an important predictor of experience-based learning [6]. This immersion potential allows decision making and taking action without any of the disruptions or influences that may affect traditional training situations [23]. Nevertheless, the comparison between the pretest and posttest showed no significant difference regarding the results of the knowledge test, nor was there a significant correlation between the experience of presence and the posttest results for the knowledge test.

In our study, we found a strong correlation between the experience of presence and situational, intrinsic motivation. On the one hand, this may be due to high covariance caused by the novelty effect. On the other hand, a more recent study assumed a strong influence of the experience of presence on affective-emotional learning goals [8]. A comparison between the pretest and posttest results showed that VR training led to a significant increase in intrinsic motivation among the participants.

The total mean score of the SUS was 65.56 points. According to the interpretation scheme of Bangor and colleagues [24], this value corresponds to a sufficient to satisfactory rating. Some users had problems particularly with real body movement, caused by the VR headset cables. Wireless VR headsets, which have become available in the meantime, will probably reduce this problem in the future. In particular, the results for the assessment of the CL dimensions showed a higher value for ECL than for ICL or GCL. This is most likely due to usability issues, which can lead to a kind of split attention effect when trainees focus their attention on both the virtual and physical worlds [25]. Makransky and Lilleholt [8] used a structural equation model to analyze the interrelationships between VR features, presence experience, usability, learning processes, and learning results. In their analysis, they were able to show that usability influences the presence experience as well as the cognitive learning processes and therefore also has indirect and direct influences on the learning results. These findings indicate which fields should be addressed in the future to achieve overall improvement of the VR training system.

Limitations

This feasibility study comprised a small sample of experienced emergency physicians who, by participating, expressed great interest in innovative training methods. A cross-sectional, one-group pretest-posttest design has many limitations regarding internal validity. As this VR simulation training was used only in the context of a first-time and one-time test phase, the results reported in this paper provide only a statement about this specific training duration.

A comparison of our results with other, similar studies appears problematic; due to the current variations in the configurations of VR technologies and methods used in studies, no comparable context of application could be referenced [26]. In addition, the use of subjective assessments as the primary outcome variable is associated with lower validity than the use of objective variables [27,28]. Furthermore, in future studies, validated assessment tools for nontechnical and clinical reasoning skills should be used instead of knowledge tests, which only measure declarative knowledge. These tools should take into consideration the specificity of VR simulation (eg, as a VR embedded assessment tool) [29].

Female participants were underrepresented in the sample, but this corresponds to the low percentage of women among German emergency physicians [30,31]. Bangor and colleagues [32] were able to show in an empirical study that there is no significant connection between gender and SUS scale values. Saredakis and colleagues [33] found no gender difference regarding VR sickness when using head-mounted displays. Nevertheless, in further studies, additional variables related to personality traits of the media users should be included in the analysis. The following variables should be considered: visual skills, cognitive skills such as spatial intelligence, openness towards new experiences, and immersive tendencies [34,35].

Conclusions

As in previous evaluation studies with paramedic trainees, experienced emergency physicians assessed the virtual simulation training similarly with regard to its perceived effectiveness. In development projects where innovative educational technologies are being designed and tested, it is important to continuously measure the specific media use factors in addition to the variables for determining training effectiveness. With respect to the further development of the VR system, usability problems or technical problems should be resolved and analyzed in a more differentiated manner. Technological factors (eg, tracking, field of view, latency) that enable interactivity of the user in immersive VR have a strong negative influence on the experience of presence, CL, and performance [36-38]. In terms of research methodology, follow-up studies with control groups and a longitudinal crossover design should be initiated.

Acknowledgments

We thank our partners in the EPICSAVE project consortium and Priv. Doz. Dr. Jan Breckwoldt, Institute of Anesthesiology, University Hospital Zurich, Switzerland, for the critical review of the manuscript. We also thank Sonnhild Namingha from Fraunhofer IESE for assistance in the translation of this paper. This work was carried out within the project EPICSAVE, which was funded by the German Federal Ministry of Education and Research and by the European Union (European Social Fund, FKZ 01PD15004).

Authors' Contributions

DL participated in study conception, data acquisition, data analysis, and writing of the manuscript. SM and MG participated in study conception and data acquisition. TL and JS participated in study conception, media design, and the writing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

DL, TL, and JS were involved in the VR software development process during the project EPICSAVE.

Multimedia Appendix 1
The knowledge test for pre- and posttest.
[DOCX File , 23 KB - games_v8i3e18822_app1.docx]

Multimedia Appendix 2
The primary data set of pretest data.
[XLSX File (Microsoft Excel File), 17 KB - games_v8i3e18822_app2.xlsx]

Multimedia Appendix 3
The primary data set of posttest data.
[XLSX File (Microsoft Excel File), 28 KB - games_v8i3e18822_app3.xlsx]

References

1. Urban B, Lazarovici M, Sandmeyer B. Simulation in der Notfallmedizin- stationäre Simulation. In: St. Pierre M, Breuer G, editors. Simulation in der Medizin: Grundlegende Konzepte - Klinische Anwendung. Heidelberg: Springer; 2018:309-335.
2. Marung H, Höhn M, Gräsner J, Adler J, Schlechtriemen T. NASIM 25 – an option to improve emergency physician training. Notfall Rettungsmed 2016 Jul 14;19(7):548-553 German. [doi: 10.1007/s10049-016-0204-3]
3. Armbruster W, Kubulus D, Schlechtriemen T, Adler J, Höhn M, Schmidt D, et al. [Improvement of emergency physician education through simulator training. Consideration on the basis of the model project "NASimSaar25"]. Anaesthesist 2014 Sep;63(8-9):691-696. [doi: 10.1007/s00101-014-2353-4] [Medline: 25056410]
4. McGrath JL, Taekman JM, Dev P, Danforth DR, Mohan D, Kman N, et al. Using Virtual Reality Simulation Environments to Assess Competence for Emergency Medicine Learners. Acad Emerg Med 2018 Feb;25(2):186-195 [FREE Full text] [doi: 10.1111/acem.13308] [Medline: 28888070]
5. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
6. Mantovani F, Castelnuovo G. Sense of Presence in Virtual Training: Enhancing Skills Acquisition and Transfer of Knowledge through Learning Experience in Virtual Environments. In: Riva G, Davide F, Ijsselsteijn W, editors. Being There - Concepts, Effects And Measurements Of User Presence In Synthetic Environments. Amsterdam, Oxford: IOS; Jan 01, 2003:167-182.
7. Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education 2020 Apr;147:103778. [doi: 10.1016/j.compedu.2019.103778]
8. Makransky G, Lilleholt L. A structural equation modeling investigation of the emotional value of immersive virtual reality in education. Education Tech Research Dev 2018 Feb 22;66(5):1141-1164. [doi: 10.1007/s11423-018-9581-2]

9. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients--what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ 2015 Feb 01;15:11 [FREE Full text] [doi: 10.1186/s12909-015-0296-3] [Medline: 25638167]
10. Lerner D, Pranghofer J, Franke A. The impact of presence experience on learning and training effects in a virtual reality simulation environment. Pädagogik der Gesundheitsberufe 2020;7(1):23-31 German. [doi: 10.24406/fordatis/6]
11. Ring J, Beyer K, Biedermann T, Bircher A, Duda D, Fischer J, et al. Guideline for acute therapy and management of anaphylaxis: S2 Guideline of the German Society for Allergology and Clinical Immunology (DGAKI), the Association of German Allergologists (AeDA), the Society of Pediatric Allergy and Environmental Medicine (GPA), the German Academy of Allergology and Environmental Medicine (DAAU), the German Professional Association of Pediatricians (BVKJ), the Austrian Society for Allergology and Immunology (ÖGAI), the Swiss Society for Allergy and Immunology (SGAI), the German Society of Anaesthesiology and Intensive Care Medicine (DGAI), the German Society of Pharmacology (DGP), the German Society for Psychosomatic Medicine (DGPM), the German Working Group of Anaphylaxis Training and Education (AGATE) and the patient organization German Allergy and Asthma Association (DAAB). Allergo J Int 2014;23(3):96-112 [FREE Full text] [doi: 10.1007/s40629-014-0009-1] [Medline: 26120521]
12. Sclar DA, Lieberman PL. Anaphylaxis: underdiagnosed, underreported, and undertreated. Am J Med 2014 Jan;127(1 Suppl):S1-S5. [doi: 10.1016/j.amjmed.2013.09.007] [Medline: 24384132]
13. Carrillo E, Hern HG, Barger J. Prehospital Administration of Epinephrine in Pediatric Anaphylaxis. Prehosp Emerg Care 2016;20(2):239-244. [doi: 10.3109/10903127.2015.1086843] [Medline: 26555274]
14. Bernhard M, Aul A, Helm M, Mutzbauer T, Kirsch J, Brenner T, et al. Invasive Notfalltechniken in der Notfallmedizin. Notfall Rettungsmed 2008 Jul 11;11(5):304-309 German. [doi: 10.1007/s10049-008-1037-5]
15. Gaba DM. The future vision of simulation in health care. Qual Saf Health Care 2004 Oct;13 Suppl 1:i2-10 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i2] [Medline: 15465951]
16. Ritzmann S, Hagemann V, Kluge A. The Training Evaluation Inventory (TEI) - Evaluation of Training Design and Measurement of Training Outcomes for Predicting Training Success. Vocations and Learning 2013 Sep 3;7(1):41-73. [doi: 10.1007/s12186-013-9106-4]
17. Krebs R. Instructions for the Development of Multiple-Choice Questions and Tests in Medical Education. https://www.iml.unibe.ch/themen/uebersichten/artikel/wie-mc-fragen-pruefungsreif-werden. Bern, Switzerland: Institute for Medical Education, University of Bern, Switzerland; 2014. URL: https://www.iml.unibe.ch/attachment/7/download/mc_anleitung.pdf [accessed 2020-07-07]
18. Schubert T, Friedmann F, Regenbrecht H. The Experience of Presence: Factor Analytic Insights. Presence: Teleoperators and Virtual Environments 2001 Jun;10(3):266-281. [doi: 10.1162/105474601300343603]
19. Brooke J. SUS. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry. London: Taylor & Francis; 1996:189-194.
20. Klepsch M, Schmitz F, Seufert T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. Front Psychol 2017;8:1997 [FREE Full text] [doi: 10.3389/fpsyg.2017.01997] [Medline: 29201011]
21. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
22. Guay F, Vallerand RJ, Blanchard C. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). Motivation and Emotion 2000 Sep;24(03):175-213 [FREE Full text] [doi: 10.1023/A:1005614228250]
23. Winn W, Jackson R. Fourteen Propositions about Educational uses of Virtual Reality. Educational Technology 1999;39(4):5-14.
24. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
25. van der Land S, Schouten AP, Feldberg F, van den Hooff B, Huysman M. Lost in space? Cognitive fit and cognitive load in 3D virtual environments. Computers in Human Behavior 2013 May;29(3):1054-1064. [doi: 10.1016/j.chb.2012.09.006]
26. Jensen L, Konradsen F. A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol 2017 Nov 25;23(4):1515-1529 [FREE Full text] [doi: 10.1007/s10639-017-9676-0]
27. von Wendt CEA, Niemi-Murola L. Simulation in Interprofessional Clinical Education: Exploring Validated Nontechnical Skills Measurement Tools. Simul Healthc 2018 Apr;13(2):131-138. [doi: 10.1097/SIH.0000000000000261] [Medline: 29117089]
28. Chauvin A, Truchot J, Bafeta A, Pateron D, Plaisance P, Yordanov Y. Randomized controlled trials of simulation-based interventions in Emergency Medicine: a methodological review. Intern Emerg Med 2018 Apr;13(3):433-444 [FREE Full text] [doi: 10.1007/s11739-017-1770-1] [Medline: 29147942]
29. Bracq M, Michinov E, Jannin P. Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals: A Systematic Review. Simul Healthc 2019 Jun;14(3):188-194 [FREE Full text] [doi: 10.1097/SIH.0000000000000347] [Medline: 30601464]

30. Luiz T, Jung J, Flick S. [Quo vadis, preclinical emergency physician?: results of a survey of emergency medical services in Rhineland-Palatinate (Germany)]. Anaesthesist 2014 Apr;63(4):294-302. [doi: 10.1007/s00101-014-2309-8] [Medline: 24595443]
31. Ilper H, Kunz T, Walcher F, Zacharowski K, Byhahn C. [An online emergency physician survey - demography, education and experience of German emergency physicians]. Dtsch Med Wochenschr 2013 Apr;138(17):880-885. [doi: 10.1055/s-0033-1343128] [Medline: 23592344]
32. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
33. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. Front Hum Neurosci 2020;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
34. Wallach HS, Safir MP, Samana R. Personality variables and presence. Virtual Reality 2009 Jul 1;14(1):3-13 [FREE Full text] [doi: 10.1007/s10055-009-0124-3]
35. Ling Y, Nefs HT, Brinkman W, Qu C, Heynderickx I. The relationship between individual characteristics and experienced presence. Computers in Human Behavior 2013 Jul;29(4):1519-1530 [FREE Full text] [doi: 10.1016/j.chb.2012.12.010]
36. Cummings JJ, Bailenson JN. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. Media Psychology 2016;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
37. Wilson ML, Beadle SC, Kinsella AJ, Mattfeld RS, Hoover AW, Muth ER. Task performance in a head-mounted display: The impacts of varying latency. Displays 2020 Jan;61:101930 [FREE Full text] [doi: 10.1016/j.displa.2019.101930]
38. Frederiksen JG, Sørensen SMD, Konge L, Svendsen MBS, Nobel-Jørgensen M, Bjerrum F, et al. Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. Surg Endosc 2020 Mar;34(3):1244-1252 [FREE Full text] [doi: 10.1007/s00464-019-06887-8] [Medline: 31172325]

Abbreviations

3D: 3-dimensional.
ABCDE: Airway, Breathing, Circulation, Disability, Exposure.
CL: cognitive load.
ECL: extraneous cognitive load.
EPICSAVE: Enhanced Paramedic Vocational Training with Serious Games and Virtual Environments.
GCL: germane cognitive load.
ICL: intrinsic cognitive load.
IPQ: Igroup Presence Questionnaire.
SAMPLER: Signs/symptoms, Allergies, Medication, Past medical history, Last meal, Events prior to incident, Risk factors.
SIMS: Situational Motivation Scale.
SUS: System Usability Scale.
TEI: Training Evaluation Inventory.
VP: virtual patient.
VR: virtual reality.

Edited by G Eysenbach; submitted 21.03.20; peer-reviewed by A Kononowicz, S Jung; comments to author 20.04.20; revised version received 28.05.20; accepted 11.06.20; published 31.07.20.

Please cite as:
Lerner D, Mohr S, Schild J, Göring M, Luiz T
An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study
JMIR Serious Games 2020;8(3):e18822
URL: https://games.jmir.org/2020/3/e18822
doi:10.2196/18822
PMID:32735548

Original Paper

The Effects of Gamification on Computerized Cognitive Training: Systematic Review and Meta-Analysis

Julie F Vermeir[1], BBehSc (Hons); Melanie J White[1], PhD; Daniel Johnson[2], PhD; Geert Crombez[3], PhD; Dimitri M L Van Ryckeghem[3,4,5], PhD

[1]School of Psychology and Counselling, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
[2]Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
[3]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
[4]Department of Clinical Psychological Science, Maastricht University, Maastricht, Netherlands
[5]Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding Author:
Julie F Vermeir, BBehSc (Hons)
School of Psychology and Counselling
Institute of Health and Biomedical Innovation
Queensland University of Technology
170 Victoria Park Road
Brisbane, 4059
Australia
Phone: 61 731384714
Fax: 61 731380486
Email: julie.vermeir@hdr.qut.edu.au

Abstract

Background: There has been a growing interest in the application of gamification (ie, the use of game elements) to computerized cognitive training. The introduction of targeted gamification features to such tasks may increase motivation and engagement as well as improve intervention effects. However, it is possible that game elements can also have adverse effects on cognitive training (eg, be a distraction), which can outweigh their potential motivational benefits. So far, little is known about the effectiveness of such applications.

Objective: This study aims to conduct a systematic review and meta-analysis to investigate the effect of gamification on process outcomes (eg, motivation) and on changes in the training domain (eg, cognition), as well as to explore the role of potential moderators.

Methods: We searched PsycINFO, Cumulative Index to Nursing and Allied Health Literature, ProQuest Psychology, Web of Science, Scopus, PubMed, Science Direct, Excerpta Medica dataBASE, Institute of Electrical and Electronics Engineers Xplore, Association for Computing Machinery, and a range of gray-area literature databases. The searches included papers published between 2008 and 2018. Meta-analyses were performed using a random-effects model.

Results: The systematic review identified 49 studies, of which 9 randomized controlled trials were included in the meta-analysis. The results of the review indicated that research in this context is still developing and lacks well-controlled empirical studies. Gamification in cognitive training is applied to a large range of age groups and audiences and is mostly delivered at a research site through computers. Rewards and feedback continue to dominate the gamification landscape, whereas social-oriented features (eg, competition) are underused. The meta-analyses showed that gamified training tasks were more motivating/engaging (Hedges $g=0.72$) and more demanding/difficult (Hedges $g=-0.52$) than non- or less-gamified tasks, whereas no effects on the training domain were found. Furthermore, no variables moderated the impact of gamified training tasks. However, meta-analytic findings were limited due to a small number of studies.

Conclusions: Overall, this review provides an overview of the existing research in the domain and provides evidence for the effectiveness of gamification in improving motivation/engagement in the context of cognitive training. We discuss the shortcomings in the current literature and provide recommendations for future research.

(JMIR Serious Games 2020;8(3):e18644) doi:10.2196/18644

KEYWORDS

gamification; cognition; health; systematic review; meta-analysis

Introduction

Background

Computer-based cognitive training typically involves systematic practice on computerized tasks designed to improve cognitive abilities, such as memory, attention, or processing speed [1]. Although reviews and meta-analyses have demonstrated that such training can be beneficial in enhancing various cognitive functions, interventions are limited by their repetitive and time-consuming character [2,3]. Indeed, most often participants are required to repeat similar training trials continuously over multiple sessions across several weeks. Tasks are perceived as monotonous and boring, which can lead to disengagement and low completion and response rates [4,5]. This is particularly an issue for cognitive research where program adherence is a pertinent problem, with dropout rates often higher than 25% [6,7]. As such, new techniques to help increase task engagement are now being developed [8,9].

There is a growing interest in the application of gamification to computerized cognitive training tasks [10]. Gamification is the use of digital game elements in nonentertainment settings to increase motivation and engagement [10,11]. According to self-determination theory (SDT) [12,13], which is one of the most established theoretical frameworks within gamification research [14], motivation is multidimensional and falls on a continuum from intrinsic motivation through extrinsic motivation to amotivation. Intrinsic motivation (performing an activity for its inherent satisfaction) is a major factor for long-term engagement and long-term behavior change [12,15], whereas extrinsic motivation (performing an activity solely for its outcome) is more useful when short-term engagement and short-term changes are required [13,15]. Gamification aims to combine both types of motivation [16] by using game elements, such as points, badges, game levels, and avatars [10,17].

As the concept of gamification is relatively new [11], empirical evidence regarding the effectiveness of gamified cognitive training tasks is still emerging. The introduction of targeted gamification features to such tasks may increase motivation and engagement as well as improve intervention effects [18,19]. However, there is also the possibility for game elements to have adverse effects on cognitive training (eg, be a distraction), which can outweigh their potential motivational benefits [8]. So far, it is not quite clear how and whether modifying these *traditional* interventions influences motivation, engagement, and training success. Rigorous evaluation of gamified cognitive training tasks is necessary to help guide the design of future tasks, applications, and randomized controlled trials (RCTs).

To our knowledge, there is no meta-analysis of the quantitative body of literature that has specifically examined the effectiveness of gamification applied to cognitive training, nor a review that has examined the quality of evidence for the efficacy of such interventions. However, there has been a systematic review exploring how gamification has been used for cognitive assessment and training purposes, which identified 33 studies [10]. Although the aforementioned review offers valuable insights into how cognitive tasks have been gamified, the quality of evidence provided by the studies was not assessed. Furthermore, the use of gamification in health interventions is a rapidly growing field, especially at a time when there is an increased usage of smart mobile technology, which makes gamified interventions potentially more accessible and appealing [17]. It is therefore timely to conduct an in-depth review of the literature and to provide a meta-analytic synthesis combining all evidence on the effectiveness of gamified cognitive training tasks. Findings from this review may have important implications for technological development, clinical practice, and future research.

Aims and Objectives

This review was divided into 2 parts. The first part was a comprehensive state-of-the-art review on gamification of cognitive training tasks with the following objectives: (1) determine the types of audiences and age groups that have been investigated; (2) determine the cognitive domains that have been targeted; (3) describe the various forms in which gamification has been applied (eg, type of game elements and theories used); and (4) establish the methodological quality of available studies. The second part is a meta-analysis of the effectiveness of gamification (assessed through RCT studies) applied to cognitive training with the following objectives: (1) assess the impact of gamification on process outcomes (ie, motivation, engagement, flow, immersion, demand, difficulty, and feasibility); (2) assess the impact of gamification on changes in the training domain (ie, cognitive process and clinical outcomes); and (3) explore possible moderators.

Methods

Protocol and Registration

This systematic review and meta-analysis were conducted and reported in accordance with the Preferred Reporting Items for Systematic and Meta-Analyses guidelines [20]. Furthermore, the protocol was preregistered in the International Prospective Register of Systematic Reviews (PROSPERO registration number CRD42018082309).

Definition and Operationalization of Key Constructs

Gamification was defined and operationalized as the use of digital game elements (eg, points, avatars) in nonentertainment settings to increase motivation and engagement [10,11]. This definition allows the differentiation of gamification from serious games [11,21]. Serious games employ full-fledged games within nongame contexts (eg, an interactive world in which players complete challenges designed to improve physical activity), whereas gamification uses elements or individual features of a game embedded in real-world contexts (eg, a mobile health application that uses points and badges to encourage physical activity). In practice, however, the actual distinction between the two can be blurry and highly subjective [21]. As a result, a cautious approach was undertaken in which edge cases were discussed among the authors (JV, DJ, and MW) and resolved by consensus.

Cognitive training was defined and operationalized as training of mental processes involving attention (eg, selective attention), memory (eg, working memory), executive functioning (eg, planning, inhibition, mental flexibility), decision making, processing speed, and perception (including visual, auditory, and tactile perception).

Search Strategy

Electronic searches were performed between February 14 and 18, 2018. The following electronic databases were included in this review, which were identified as relevant to psychology, health, social science, and information technology: (1) PsycINFO (via EBSCOhost); (2) Cumulative Index to Nursing and Allied Health Literature (CINAHL; via EBSCOhost); (3) ProQuest Psychology; (4) Web of Science; (5) Scopus; (6) PubMed; (7) Science Direct; (8) Excerpta Medica dataBASE (EMBASE); (9) Institute of Electrical and Electronics Engineers (IEEE) Xplore; (10) Association for Computing Machinery (ACM); and (11) a range of gray-area literature databases. A complementary manual search of the reference lists of eligible records and relevant published reviews was also conducted to locate studies not identified in the database searches.

The search terms that we used included terms for gamification in conjunction with terms for cognitive training. Given that gamification was defined and operationalized as the use of digital game elements in nonentertainment settings to increase motivation and engagement [10,11], it was decided to include the term (and variants of the term) *game elements* in the search. This was necessary to capture research that identifies with the literature of gamification but labels it as serious game or games with a purpose rather than gamification. For each database, full and truncated search terms relating to gamification (eg, gamif*, game element*), cognition (eg, cognit*, mental process*), and training (eg, train*, intervention*) were used (see Multimedia Appendix 1 for further details on the search strategy). No limiters were applied at this stage.

Selection Criteria

For the systematic review, studies were selected if they met the following inclusion criteria:

1. Original empirical research that explicitly refers to gamification, the gamification literature, or the use of game elements. This does not include review papers such as narrative reviews, systematic reviews, or meta-analyses.
2. Peer-reviewed documents (eg, published papers, doctoral theses, study protocols, conference papers).
3. The full text is available in published or unpublished form. This does not include extended abstracts, tutorials, posters, editorials, and letters.
4. A task specifically designed to train or modify cognition as defined earlier. This includes the full range of intervention contexts (eg, health, education, rehabilitation).
5. Cognitive training tasks for which at least one game element (eg, points, avatars) has been added. This does not include studies that reported on persuasive games, serious games, or full-fledged games (eg, video games).
6. Cognitive training tasks are delivered via a digital device (eg, personal computers, laptops, tablets, smartphones, virtual reality headsets).
7. Studies report on at least one outcome related to process (eg, motivation, engagement) or to changes in the training domain (eg, cognition, affect), measured through self-report (eg, questionnaire) and/or objective (eg, dropout rate) measurement methods.
8. The full text is available in English (though research may not necessarily be conducted in English).

For meta-analysis, some additional inclusion criteria were formulated to ensure that studies have a minimal methodological quality. The studies included in the meta-analysis are a subset of the studies included in the systematic review. Additional inclusion criteria for the meta-analysis were as follows:

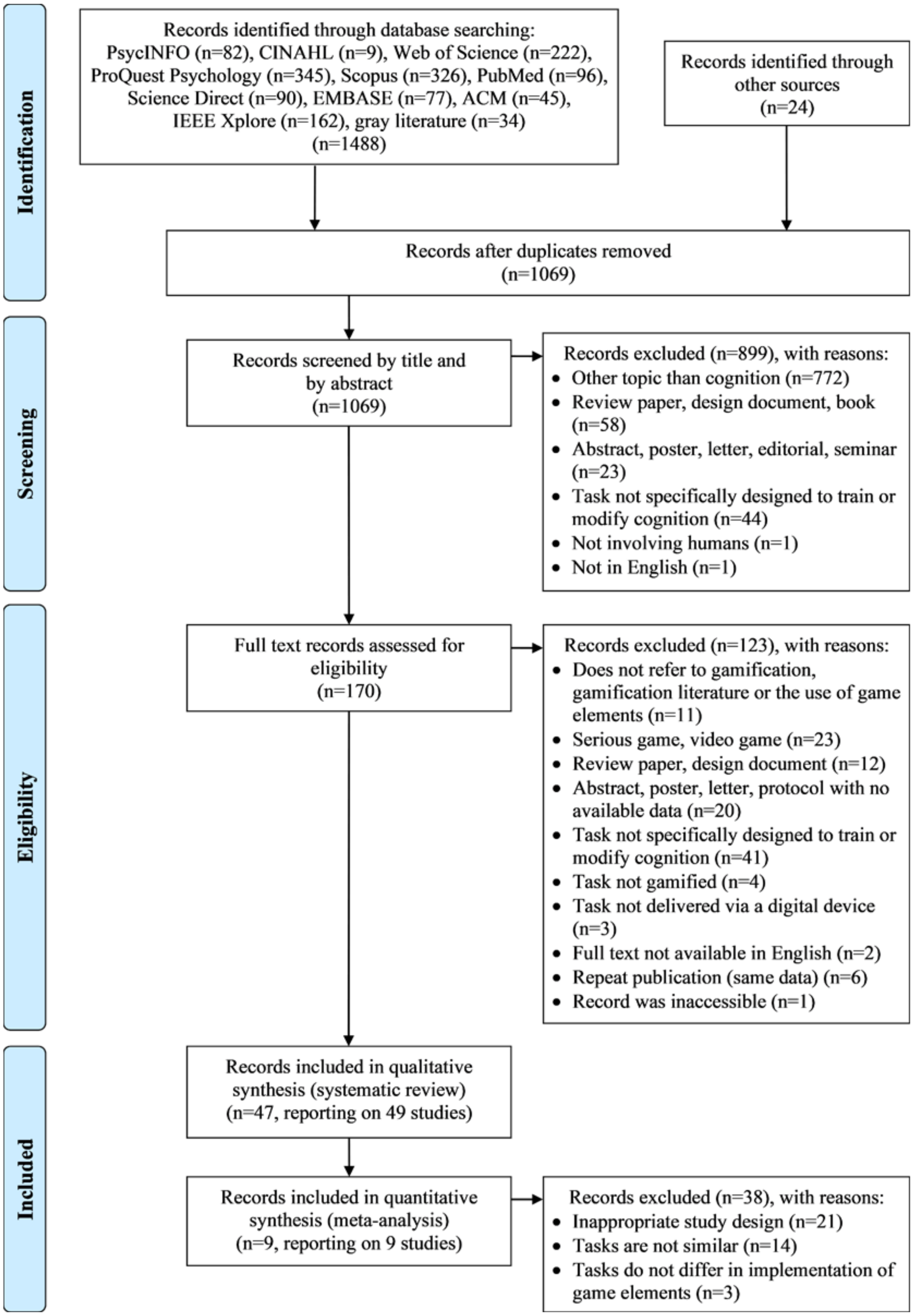
1. Studies report the effect of (at least) two cognitive training tasks that are similar to each other except for the implementation of gamification features.
2. The study design consisted of an RCT in which 2 groups performed a cognitive training task. Studies were excluded if they only reported a single-group pre-post design.

Study Selection

Owing to the lack of consensus regarding the definition of gamification, we erred on the side of caution and did not exclude documents based on whether they used terms such as persuasive games, serious games, video games, computer-based, and games with a purpose. Records were selected if they were considered relevant based on our inclusion criteria and on whether at least two authors reached consensus for inclusion.

Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart for the selection of included records in the systematic review and meta-analysis. After merging search results across databases (N=1488) using Endnote X8 citation management software, the first author (JV) removed duplicates and screened the remaining records (n=1069) by both title and abstract according to the prespecified eligibility criteria. If it was not possible to determine the eligibility of a record from the title and abstract, the full-text record was obtained. Full-text records (n=170) were retrieved and evaluated against the inclusion criteria (JV). To check the reliability of the process, a second author (MW, DJ, or DV) assessed 20% of the selected full-text records, which resulted in no disagreement. A total of 47 records, reporting on 49 independent studies, were included in the systematic review (Part 1). Of these, 9 records, reporting on 9 independent studies, were suitable for a meta-analysis (Part 2). The list of eligible records was sent to experts in the field of cognitive psychology who were asked to identify further eligible studies, and no relevant records were added. Review authors were not blinded to the authorship, institution, journal, or results.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of the study selection process.



Data Extraction Strategy

Data across all included studies were extracted into a prepared Microsoft Access form, which was developed specifically for this review. The data extraction form was piloted on a random sample of studies before being finalized. Two authors (JV and DV) independently extracted and coded the data from the included studies. Any disagreement or ambiguity was resolved by discussion and consensus, and when necessary, a third reviewer (MW, DJ, or GC) was consulted. For the meta-analysis, the corresponding author of each publication was contacted via email for further information when data were missing or unclear. In addition, the authors were invited to comment on the coding and data extraction of their study. When the requested information was not provided, or could not be provided, this was coded as *incomplete data*, and this study was not included in the meta-analyses incorporating this variable. After coding, outcome data for the meta-analysis were extracted and inputted into the Comprehensive Meta-Analysis (CMA) software, version 3.3.070 (Biostat Inc).

Coding System and Coding Decisions

Every study was coded in terms of study and sample characteristics, intervention characteristics, outcome variables, and methodological quality. For the meta-analysis, additional data were extracted, and several potentially relevant moderating variables were coded.

Study and Sample Characteristics

For study characteristics, we provided each study and each outcome within the study with an identification number (eg, study ID1.1, outcome ID1). We then coded the publication channel (published article, doctoral thesis, or conference paper), year of publication, first author's country of publication, experimental design (between-subject design, single-group design, or case study design), and the number and description of groups. Additional information, such as bibliographic reference and email address of the corresponding author, was also extracted.

For sample characteristics, we coded the overall sample size, the mean age of the sample, proportion of females, and type of participants (eg, trait-anxious adults). For each study included in the meta-analysis, we additionally coded the sample size, mean age of the participants, proportion of females, and sample size used for analyses (ie, n for each outcome) in each condition.

Intervention Characteristics

For intervention characteristics, we coded the domain of cognitive training (eg, memory), name of the training task (eg, n-back task), name of the gamified training task (eg, Shots Game), study training length, follow-up(s) after training, number of training sessions, session frequency, average session length (minutes), number of trials per session, type of stimuli used, task delivery location (eg, home), the modality of delivery (eg, tablet), and theory used to apply gamification (eg, SDT). Game elements were coded using a combination of the systems provided by Johnson et al [17], Lumsden et al [10], and Sardi et al [22] (Textbox 1).

Textbox 1. Game elements (adapted from Johnson et al [17], Lumsden et al [10], and Sardi et al [22]) and their descriptions.

Avatar

Visual representation of the user

Challenge

Users are required to overcome a challenge by introducing some pressures (eg, time limit, lives) to keep them interested in the gamified system

Competition

Users compete against other users. This can be achieved by using, for example, leaderboards

Difficulty adjustment

Difficulty is adjusted to the users' ability either automatically (known as dynamic difficulty adjustment) or manually (eg, users can select the level of difficulty)

Feedback loops

Composed of 2 or more steps. Users perform an action, receive gamified/playful feedback about their performance, and with this information can modify their behavior. Feedback loops can be positive (ie, amplifies an action) or negative (ie, reduces an action), and feedback can be provided at different timings (immediate or delayed) and through different delivery methods (eg, visual, auditory)

Levels

Visual indicators that inform users about their completed intermediate goals

Progress (task-related)

Visual features that inform users about their progression throughout the gamified system (eg, progress bar, a graphical completion chart)

Rewards

Indicators such as points (a numerical measure that quantifies users' participation and performance), badges (visual token of achievements), and other digital rewards (eg, coins, virtual money) are allocated to users for accomplishing certain tasks

Social interaction

Users interact with other users

Sound effects

Noises used to enhance the experience (eg, music background) or to reinforce a specific action (eg, buzz sound for incorrect response)

Story/theme

Graphics or text linked together within the gamified system to enhance the experience or to give meaning to tasks (also known as meaningful stories)

Outcome Variables

For the systematic review, study outcome(s) and measurement instruments used for assessing these outcomes were extracted for each study. For the meta-analysis, some additional data extraction and coding decisions were made. In particular, we decided to take a more refined approach to the categorization of outcomes stated in the study protocol (CRD42018082309). Process-related outcomes were categorized as motivation/engagement (including interest and enjoyment), flow/immersion, demand/difficulty, and feasibility (including ease of use and task clarity). Training domain–related outcomes were categorized as a cognitive process (eg, alcohol attentional bias, working memory) and clinical outcomes (eg, alcohol use, psychiatric symptoms). This final categorization was based on consensus discussions among the authors of this paper, who have research expertise in the field of gamification and cognitive psychology. To calculate effect sizes for each of these outcomes, we opted to use only the posttraining data (ie, means, standard deviations) rather than change scores (ie, the difference between pre- and posttraining data). This approach reduces the risk of selective reporting [23,24]. Next, categorical process variables having more than two categories [25] were recoded into dummy variables, with 0 as the less desirable outcome for gamification and 1 as the more desirable outcome for gamification. Finally, when a study reported results of subgroups instead of the overall results for a group of interest, we calculated the combined effect size using the available formula in the Cochrane Handbook [23] and imputed the pooled mean and standard deviation in the meta-analysis.

Quality Assessment of Included Studies

For the systematic review (Part 1), we assessed the methodological quality of each study using a quality assessment tool specifically designed for this review. Judgments of external validity were based on the description of the setting and/or location of the study and the description of the gamified task, both coded either as *adequate* or *not adequate*. The first item was coded as *adequate* if the study provided descriptions of the following features: (1) setting where participants were recruited (eg, general population); (2) type of participants (eg, undergraduates of university); and (3) location where the study took place (eg, laboratory). The second item was coded as *adequate* if the study provided at least a description of the following features: (1) the number of trials or duration of the task; (2) type of response required during the task (eg, indicate whether the arrow is pointing to the left or the right); and (3) game elements used (described using a picture or text). The criteria for internal validity were related to the description of blinding of participants (blinding of participants, no blinding of participants, or unclear) and personnel (blinding of personnel, no blinding of personnel, or unclear), and the inclusion of a comparison control condition (yes or no). Judgments were made by 2 authors (JV and DV) independently, and a consensus was reached for existing discrepancies through discussion.

For the meta-analysis (Part 2), we used the Cochrane Collaboration risk of bias tool as described in the Cochrane Handbook for Systematic Reviews of Interventions [26]. The tool covers 5 biases: selection bias (random sequence generation and allocation concealment), performance bias (blinding of participants and personnel), detection bias for each outcome separately (blinding of outcome assessment), attrition bias (incomplete outcome data), and selective outcome reporting bias. The risk of bias for each domain was then judged as either *low risk*, *high risk*, or *unclear risk*. Judgments were made by 2 authors (JV and DV) independently, using the criteria for assessing the potential risk of bias [26]. The consensus was reached for all discrepancies through discussion.

Coding of Moderators (Meta-Analysis)

The present meta-analysis examined several samples, intervention, and study characteristics as potential moderators of the effects of gamification on outcomes related to process and changes in the training domain. Specifically, because research has shown that there may be age differences in perceived benefits from the use of gamification [27], we investigated age groups (children and adolescents: mean age <18 years or adults: mean age ≥18 years) as a potential moderator. In addition, we investigated cognitive domain targeted and population targeted (low-risk group or high-risk group) as potential moderators. Studies coded with participants in the high-risk group include at-risk participants such as the elderly, people with specific health issues (eg, patients with gliomas), or people selected because of greater health risks (eg, heavily drinking students), whereas studies coded with participants in the low-risk group included participants who were not preselected due to health risks (eg, university students, primary school children). With regard to study characteristics, we examined the number of training sessions. The question can be raised whether one session is enough to see changes in the training domain targeted. Gamification can add complexity to the training environment, and game elements may initially distract participants from the core task [8]. Thus, it is possible that training effects occur only after multiple training sessions in which participants become familiar with the gamified environment. In contrast, motivation and excitement may be high in the first few sessions but decline over time when novelty decreases [22].

Finally, we explored gamification elements (ie, number and type of game elements) as potential moderators. Similar to Koivisto and Hamari [28], game elements were grouped based on their type into the following: achievement and progression-oriented (rewards, challenge, game level, progress, difficulty adjustment, feedback loops); social-oriented (competition, social interaction); or immersion-oriented (avatar, story/theme, sound effects). This decision was made due to studies using different combinations of game elements, making the analysis of individual game elements unfeasible.

Meta-Analytic Procedures

To investigate the effectiveness of gamification applied to cognitive training tasks, 2 separate sets of meta-analyses were conducted. The first set assessed the impact of gamification on process-related outcomes (ie, motivation, engagement, flow, immersion, demand, difficulty, and feasibility) and variables moderating this relationship. The second set of meta-analyses explored the effects of gamification on changes in the training domain (ie, cognitive process and clinical outcomes) and variables moderating this relationship.

A random-effects model was chosen to combine effect sizes from different studies. We used Hedges g, which corrects for small sample sizes, to calculate effect sizes for outcome data based on posttraining means and standard deviations [29]. A negative or positive Hedges g (significant at $P<.05$) indicated that the gamified condition performed more poorly or better than the control condition (ie, a non- or less- gamified version), respectively. The degree of heterogeneity in effect sizes (significant at $P<.05$) was assessed using the Cochran Q test [30].

The sources of heterogeneity were explored by conducting subgroup and meta-regression analyses. For moderator analyses with categorical coded variables (eg, age group), we used a mixed-effects model and opted to pool within-group estimates of Tau-squared due to the limited number of studies per condition. Heterogeneity within each group (Q_w) and heterogeneity between groups (Q_b) were also evaluated. A significant Q_b ($P<.05$) indicates a significant difference in the magnitude of the effect sizes between categories of the moderator variable. If there were less than three studies within a subgroup, moderator analyses were not conducted. To maintain the independence of data, whenever necessary, effect sizes were averaged across different outcomes. For the coded continuous variables (eg, number of training sessions), we performed meta-regressions using the method-of-moments procedure with the Knapp-Hartung correction, where the slope (β) and its *P* value indicated the importance of this moderator in understanding linear changes in effect sizes [31].

Publication bias was assessed by visual inspection of funnel plots and the Egger regression test, with asymmetric funnel plots and the significant Egger test ($P<.05$) indicating publication bias [32]. When the funnel plot inspection or Egger test suggested the presence of publication bias, the trim-and-fill procedure by Duval and Tweedie [33] was applied. This procedure provides an adjusted estimate of effect size after accounting for publication bias. Publication bias analyses were conducted if at least three studies were available in the overall meta-analyses. Finally, a leave-one-out sensitivity analysis was performed by iteratively removing one study at a time to confirm that our findings were not driven by any single study. All analyses were performed using the CMA software. The magnitude of the effect sizes was categorized as small (≥0.20 to <0.50), moderate (≥0.50 to <0.80), or large (≥0.80) [34].

Results

Systematic Review of the Gamification of Cognitive Training (Part 1)

Summary of Included Studies

The key characteristics of the 49 studies included in the systematic review are summarized in Multimedia Appendix 2. A total of 78% (n=38/49) of articles were published in academic journals, 18% (n=9/49) were conference papers, and 4% (n=2/49) were doctoral theses. Studies were published between 2008 and 2017, with the majority (n=46/49, 94%) published after 2010, indicating that research on gamified cognitive training tasks is a young but rapidly growing field of research. Most of the studies (n=32/49, 65%) were undertaken in European countries, 25% (n=12/49) in the United States, 4% (n=2/49) in Australia, 4% (n=2/49) in Asian countries, and 2% (n=1/49) in Canada.

The included studies involved a total of 4003 participants. Sample sizes ranged from 3 [35,36] to 794 [37] participants. Notably, most studies (n=47/49, 96%) had fewer than 200 participants in total (often split over different conditions). When reported (n=44/49, 90%), the mean age of the participants varied from 8.98 [38] to 82.70 years [39], showing that gamification has been applied to a large range of age groups. The majority of the studied samples (n=43/49, 88%) included both males and females, in which the proportion of females varied from 9% [40] to 93% [39]. Two studies focused on male-only samples [35,41], 1 study only included female participants [42], and 3 studies did not specify the gender ratio [43-45]. Over three-fourth (n=38/49, 78%) of the studies were conducted with adults, whereas the remaining one-fourth involved children and adolescents. Of the 49 studies, 15 (31%) were delivered to low-risk samples (eg, university students, schoolchildren), 33 (67%) to high-risk groups (eg, elderly drivers, children with autism spectrum disorder), and 1 study involved both populations [46].

A range of study designs were represented across the 49 studies, with the majority (n=40/49, 82%) using a between-group (pre-post or post only) design with 2, 3, 4, 6, or 7 groups. Other study designs consisted of single-group designs (n=7/49, 14%) and case study designs (n=2/49, 4%). A total of 31 (63%) studies conducted single-domain cognitive training, with attention (n=11/49, 22%) and working memory (n=9/49, 18%) being most often targeted. The remaining studies trained multiple domains of cognition simultaneously. With respect to the reported outcomes, research on gamified cognitive training tasks is diverse. Some studies explored the effects of gamification on process-related outcomes (eg, motivation, engagement, feasibility), whereas others investigated changes in the training domain (eg, cognition, anxiety, attention-deficit/hyperactivity disorder [ADHD] symptoms); others reported a mixture of outcomes. There was also a broad variety of outcome measures used across the studies (eg, questionnaires, semistructured interviews, video-recorded observations). The final coding scheme for outcome measures is available from the authors upon request.

Overall, interventions ranged in length from one session (range 10-80 min) to multiple sessions (maximum 59 sessions; range 5-72 min), lasting up to 18 months [37]. All studies evaluated outcomes at the end of the training, but only about one-third (n=14/49, 29%) reported additional follow-up assessments, which ranged from 1 week ([47], study 1) to 5 years [48]. When the study location was provided (n=42/49, 86%), interventions were delivered at a research site (eg, laboratory, hospital, school; n=19/49, 39%), at participants' homes (n=10/49, 20%), web-based (n=5/49, 10%), at a location of choice (n=1/49, 2%), or a mixture thereof (eg, laboratory and/or home; n=7/49, 14%). The chief modalities employed for delivering gamified cognitive training were the following: (1) computers (ie, desktop, laptop, and notebook; n=26/49, 53%); (2) tablets (n=13/49, 27%); (3) smartphones (n=2/49, 4%); (4) and iPod touch (n=2/49, 4%). Several studies (n=5/49, 10%) used a combination of devices (eg, computer or tablet), and one study did not provide information on the mode of delivery [49].

The 49 studies included in this systematic review employed a range of game elements embedded in a variety of ways. All studies incorporated a combination of game elements into each intervention (range 2-9) and the mode count of gamification features was 5 (Multimedia Appendix 2). The most frequently used game elements were rewards (eg, points, badges), followed by feedback loops, story/theme, and difficulty adjustment (Table 1). Most of these game elements can be categorized as achievement and progression-oriented game features, which form the most common category of game elements used in cognitive training. Interestingly, only around one-third (n=17/49, 35%) of the studies draw upon existing theories and principles for designing gamified interventions; specifically, SDT [12,13] (n=3/49, 6%), flow theory [50] (n=2/49, 4%), the framework by Kiili and Perttula [51] (n=2/49, 4%), operant conditioning (n=1/49, 2%), principles by Gee [52] (n=1/49, 2%), or a combination of several theories and principles (n=8/49, 16%).

Table 1. Type of game elements used in the selected studies.

Game elements	Count, n
Achievement and progression-oriented	
Rewards	40
Feedback loops	39
Difficulty adjustment (dynamic and/or manual)	34
Game level	26
Progress	25
Challenge	24
Social-oriented	
Competition	3
Social interaction	2
Immersion-oriented	
Story/theme	38
Sound effects	19
Avatar	16

Quality of Included Studies

The ratings of methodological quality for studies included in the systematic review are provided in Multimedia Appendix 2. Overall, all studies (with the exception of that of Mohammed et al [53]) had at least one of the applicable criteria not fulfilled. Regarding external validity, most studies adequately described the setting and/or location of the study (n=40/49, 82%) and the gamified task (n=43/49, 88%). However, for the internal validity criteria, most studies had lower quality assessment ratings due to inadequate or unclear blinding of participants (n=38/49, 78%) and personnel (n=39/49, 80%), and the absence of a control (ie, a non- or less- gamified version) comparison condition (n=40/49, 82%).

Meta-Analysis of the Effectiveness of Gamified Cognitive Training (Part 2)

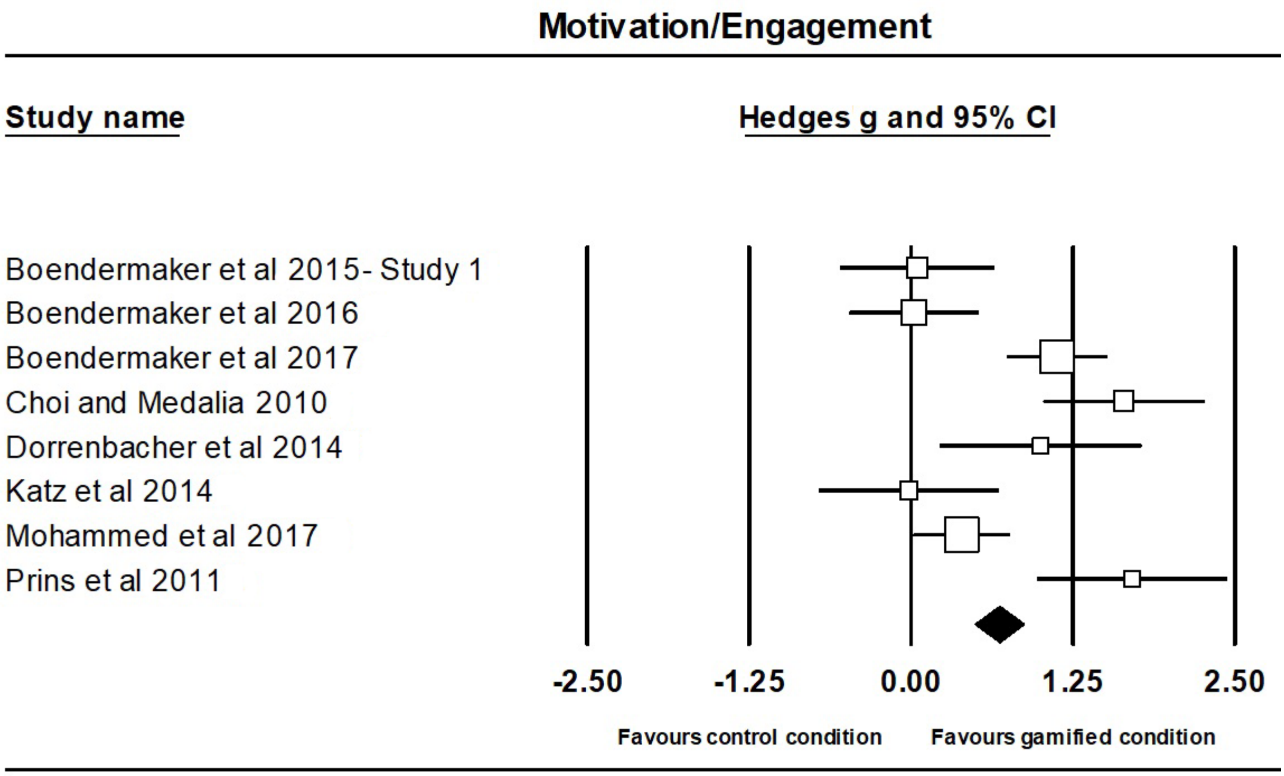
Effects of Gamification on Process-Related Outcomes

Motivation/Engagement

The meta-analysis on the effect of gamification on motivation/engagement-related outcomes contained 8 independent studies, involving a total of 514 participants, with 295 participants (M_{age}=18.88; $M_{\%females}$=48) in the gamified groups and 219 participants (M_{age}=19.23; $M_{\%females}$=56) in the comparison groups (ie, non- or less-gamified version).

The effect of gamification on motivation/engagement was moderate, positive, and significant (Hedges g=0.72; 95% CI 0.26 to 1.19; *P*=.002), indicating that gamified tasks were more motivating/engaging than non- or less-gamified tasks (Figure 2). There was significant heterogeneity between studies (Q_7=39.73; *P*<.001), warranting the relevance of moderator analyses (see section Subgroup and Meta-Regression Analyses). A sensitivity analysis revealed that the results were not significantly altered by the removal of any one study, indicating the reliability of the findings (Multimedia Appendix 3). A visual inspection of the funnel plot (Multimedia Appendix 3) and Egger test (t_6=0.37; *P*=.72) showed that there was no evidence of publication bias.

Figure 2. Forest plot of overall effect sizes comparing gamified condition and control condition (ie, non- or less-gamified version) on motivation/engagement for individual studies in alphabetical order.

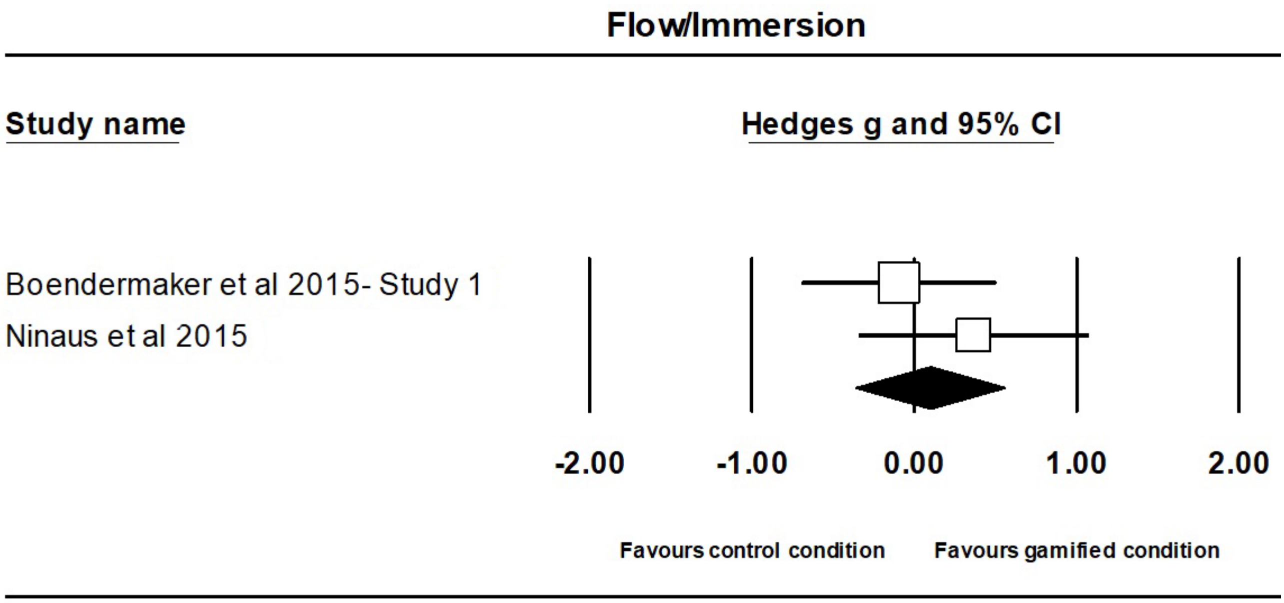


Flow/Immersion

The meta-analysis on the effect of gamification on flow/immersion-related outcomes contained 2 independent studies, involving a total of 79 participants, with 47 participants (M_{age}=22.85; $M_{\%females}$=49) in the gamified groups and 32 participants (M_{age}=23.69; $M_{\%females}$=72%) in the comparison groups (ie, non- or less-gamified version).

The effect of gamification on flow/immersion was not significant (Hedges g=0.10; 95% CI 0.36 to 0.55; *P*=.68; Figure 3), with no evidence of significant heterogeneity (Q_1=0.95; *P*=.33). There have been too few studies to allow the assessment of publication bias and to perform a sensitivity analysis.

Figure 3. Forest plot of overall effect sizes comparing gamified condition and control condition (ie, non- or less-gamified version) on flow/immersion for individual studies in alphabetical order.

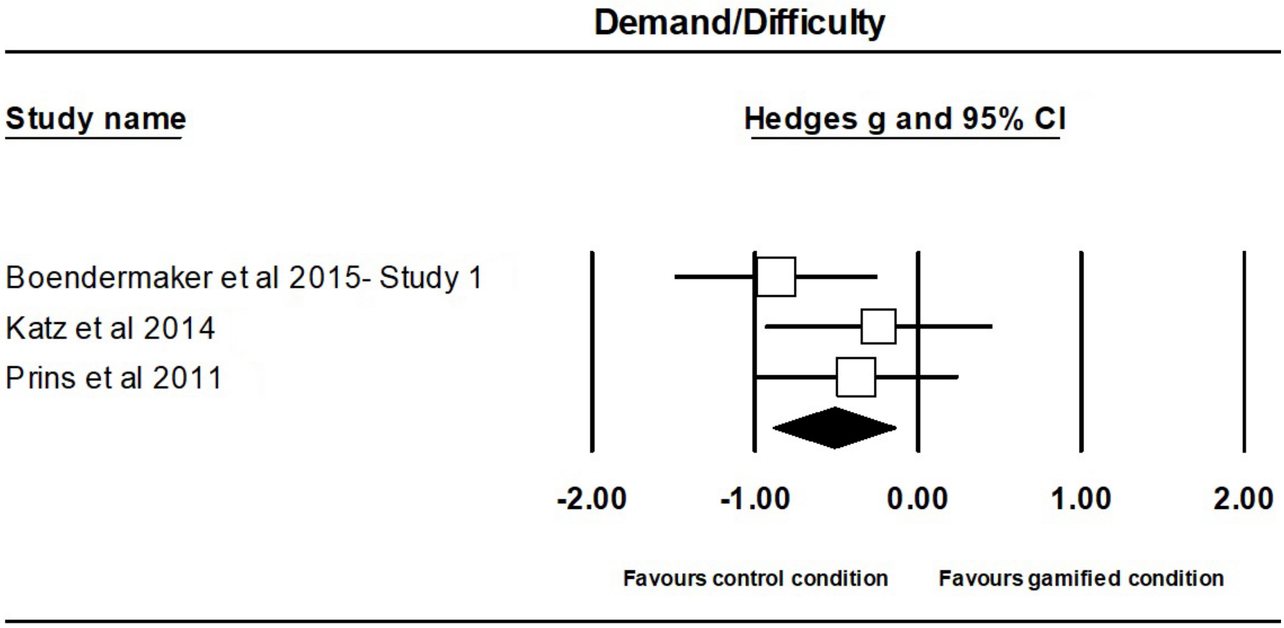


Demand/Difficulty

The meta-analysis on the effect of gamification on demand/difficulty-related outcomes contained 3 independent studies, involving a total of 136 participants, with 86 participants (M_{age}=15.24; $M_{\%females}$=35) in the gamified groups and 50 participants (M_{age}=13.60; $M_{\%females}$=26) in the comparison groups (ie, non- or less-gamified version).

The effect of gamification on task demand/difficulty was moderate, negative, and significant (Hedges g=0.52; 95% CI 0.89 to 0.14; *P*=.007), indicating that gamified tasks were more demanding/difficult than non- or less-gamified tasks (Figure 4). There was no evidence of significant heterogeneity between studies (Q_2=2.03; *P*=.36). A sensitivity analysis showed significant changes in the results, possibly because of the low statistical power of the analysis of 2 studies only (Multimedia Appendix 3). A visual inspection of the funnel plot (Multimedia Appendix 3) and Egger test (t_1=0.91; *P*=.53) showed that there was no evidence of publication bias.

Figure 4. Forest plot of overall effect sizes comparing gamified condition and control condition (ie, non- or less-gamified version) on demand/difficulty for individual studies in alphabetical order.



Feasibility-Related Outcomes

The effect of gamification on feasibility was only reported in 1 study ([47], study 1). The results of this study indicated that gamified tasks (n=34; M_{age}=22.64; $M_{\%females}$=44) were less easy to use or less clear than the control task (n=15; M_{age}=23.20; $M_{\%females}$=47); however, this difference was not significant (Hedges g =−0.36; 95% CI −0.97 to 0.26; *P*=.25). There have been too few studies to perform any further analyses.

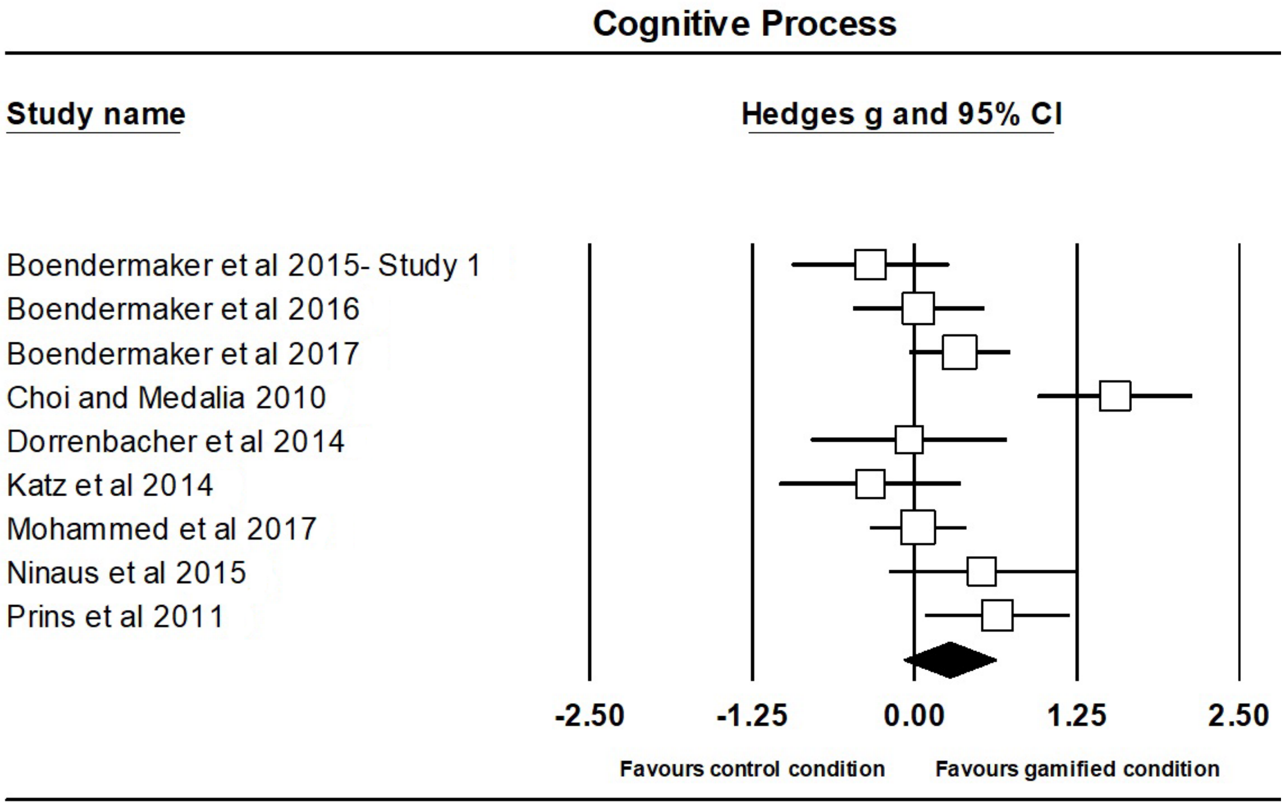
Effects of Gamification on Training Domain–Related Outcomes

Cognitive Process

The meta-analysis on the effect of gamification on cognitive process outcomes contained 9 independent studies, involving a total of 530 participants, with 301 participants (M_{age}=19.10; $M_{\%females}$=49) in the gamified groups and 229 participants (M_{age}=9.73; $M_{\%females}$=59) in the comparison groups (ie, non- or less-gamified version).

The effect of gamification on cognitive process outcomes was small and positive, but not significant (Hedges g=0.27; 95% CI 0.08 to 0.62; *P*=.14; Figure 5). There was significant heterogeneity between studies (Q_8=29.91; *P*<.001), warranting the relevance of moderator analyses (see section Subgroup and Meta-Regression Analyses). A sensitivity analysis revealed that the results were not significantly altered by the removal of any one study (Multimedia Appendix 3). Visual inspection of the funnel plot (Multimedia Appendix 3) and Egger test (t_7=0.15; *P*=.88) showed that there was no evidence of publication bias.

Figure 5. Forest plot of overall effect sizes comparing gamified condition and control condition (ie, non- or less-gamified version) on the cognitive process for individual studies in alphabetical order.

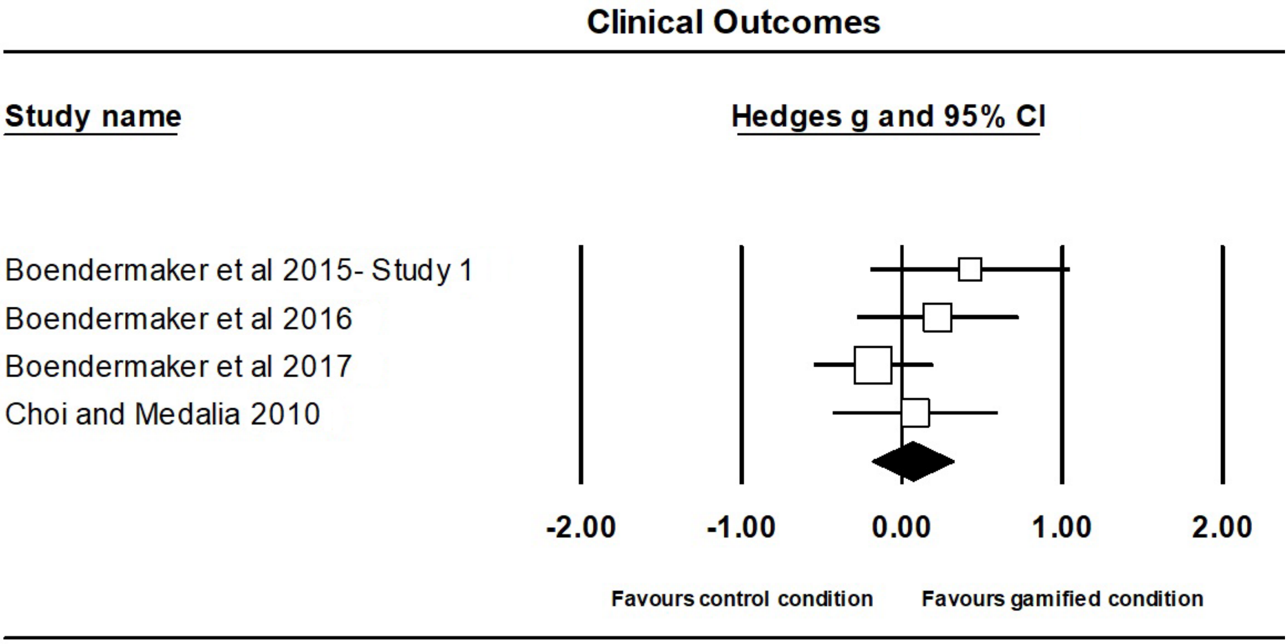


Clinical Outcomes

The meta-analysis on the effect of gamification on clinical outcomes contained 4 independent studies, involving a total of 276 participants, with 158 participants (M_{age}=21.88; $M_{\%females}$=2) in the gamified groups and 119 participants (M_{age}=22.94; $M_{\%females}$=65) in the comparison groups (ie, non- or less-gamified version).

The effect of gamification on clinical outcomes was not significant (Hedges g=0.07; 95% CI 0.19 to 0.32; *P*=.61; Figure 6), with no evidence of significant heterogeneity (Q_3=3.32; *P*=.35). A sensitivity analysis revealed that the results were not significantly altered by the removal of any 1 study (Multimedia Appendix 3). A visual inspection of the funnel plot (Multimedia Appendix 3) and Egger test (t_2=5.59; *P*=.03) showed that there was evidence of publication bias. The analyses undertaken using the trim-and-fill approach by Duval and Tweedie [33] did not change the overall effect (Hedges g=0.08; 95% CI 0.36 to 0.20; 2 studies were trimmed). Follow-up effect sizes for clinical outcomes were not calculated because only 1 study reported follow-up [54].

Figure 6. Forest plot of overall effect sizes comparing gamified condition and control condition (ie, non- or less-gamified version) on clinical outcomes for individual studies in alphabetical order.



Subgroup and Meta-Regression Analyses

To test possible explanations of observed heterogeneity for motivation/engagement and cognitive process outcomes, several subgroup analyses and meta-regressions were conducted.

Age Group

Subgroup analyses were performed to examine whether effect sizes varied according to age category (children and adolescents vs adults). There was no significant difference in effect sizes between studies targeting children and adolescents and studies targeting adults for motivation/engagement outcomes (Q_1=0.94; *P*=.33) and cognitive process outcomes (Q_1=0.16; *P*=.69; Tables 2 and 3).

Table 2. Results of the subgroup analyses for motivation/engagement outcomes.

Moderator	n[a]	k[b]	g[c]	95% CI	*P* value	Q_w[d]	*P* value	Q_b[e]	*P* value
Cognitive domain								—[f]	—
Arithmetic ability	57	1	1.64	0.13 to 3.16	.03	—	—		
Attention	61	1	0.02	-1.45 to 1.49	.98	—	—		
Executive control	27	1	1.00	-0.59 to 2.59	.22	—	—		
Inhibition	169	2	0.62	-0.43 to 1.65	.25	8.84	.003		
Working memory	200	3	0.67	-0.20 to 1.54	.13	12.74	.002		
Age group								0.94	.33
Adults	282	4	0.51	-0.12 to 1.13	.11	18.86	<.001		
Children and adolescents	232	4	0.96	0.30 to 1.61	.004	12.19	.007		
Population type								0.14	.71
Low risk	298	4	0.63	-0.08 to 1.34	.08	11.99	.007		
High risk	216	4	0.82	0.10 to 1.54	.03	27.74	<.001		
Game elements type								—	—
Achievement, progression, and immersion	408	6	0.69	0.12 to 1.26	.02	26.12	<.001		
Achievement, progression, immersion, and social	106	2	0.84	-0.15 to 1.83	.10	13.14	<.001		

[a]n: combined sample size.

[b]k*:* number of studies.

[c]g: Hedges g.

[d]Q_w: heterogeneity statistics within each group.

[e]Q_b: heterogeneity statistics between groups.

[f]Not available due to insufficient observations.

Table 3. Results of the subgroup analyses for cognitive process outcomes.

Moderator	n [a]	k [b]	g [c]	95% CI	*P* value	Q_w [d]	*P* value	Q_b [e]	*P* value
Cognitive domain								—[f]	—
Arithmetic ability	57	1	1.54	0.66 to 2.43	.001	—	—		
Attention	61	1	0.03	–0.80 to 0.86	.94	—	—		
Executive control	26	1	–0.04	–1.04 to 0.96	.93	—	—		
Inhibition	155	2	0.06	–0.52 to 0.64	.85	3.53	.06		
Working memory	231	4	0.21	–0.23 to 0.64	.35	6.28	.10		
Age group								0.16	.69
Adults	310	5	0.34	–0.17 to 0.85	.20	24.46	<.001		
Children and adolescents	220	4	0.18	–0.40 to 0.76	.54	5.44	.14		
Population type								0.89	.35
Low risk	314	5	0.11	–0.38 to 0.60	.66	4.60	.33		
High risk	216	4	0.46	–0.08 to 1.00	.10	22.58	< .001		
Game elements type								—	—
Achievement, progression, and immersion	425	7	0.18	–0.22 to 0.58	.39	7.66	.26		
Achievement, progression, immersion, and social	105	2	0.61	–0.16 to 1.37	.12	18.94	< .001		

[a] *n*: combined sample size.

[b] *k:* number of studies.

[c] *g*: Hedges *g*.

[d] $Q_{w:}$ heterogeneity statistics within each group.

[e] $Q_{b:}$ heterogeneity statistics between groups.

[f] Not available due to insufficient observations.

Population Type

Subgroup analyses were performed to examine whether effect sizes varied according to the population type (low-risk vs high-risk groups). There was no significant difference in effect sizes between training tasks targeting low-risk samples (eg, university students) and tasks targeting high-risk groups (eg, elderly drivers) for motivation/engagement outcomes (Q_1=0.14; *P*=.71) and cognitive process outcomes (Q_1=0.89; *P*=.35; Tables 2 and 3).

Cognitive Domain

Subgroup analyses according to the cognitive domain targeted were not possible because there were fewer than 3 studies within some subgroups for both motivation/engagement and cognitive process outcomes (Tables 2 and 3).

Game Elements Type

Subgroup analyses according to the type of game elements used were not possible because there were fewer than 3 studies within a subgroup for both motivation/engagement and cognitive process outcomes (Tables 2 and 3). It is noteworthy that, based upon categorization, we were unable to investigate the impact of a single category (eg, achievement and progression features) because studies always incorporated a combination of categories into their tasks. In particular, studies reported only outcomes from the following combined categories: (1) achievement, progression, and immersion-oriented game features and (2) achievement, progression, immersion, and social-oriented game features (Tables 2 and 3).

Number of Game Elements

Meta-regression analyses were performed to examine whether the effects of gamification on outcomes of interest varied according to the number of game features used (independent of the game element type). There was no significant relationship between the number of game elements used and effect sizes (Hedge g) for motivation/engagement outcomes (point estimate of slope=0.26; 95% CI 0.57 to 0.06; *P*=.09; Multimedia Appendix 3) and cognitive process outcomes (point estimate of slope=0.21; 95% CI 0.46 to 0.04; *P*=.08; Multimedia Appendix 3).

Number of Training Sessions

Meta-regression analyses were performed to examine whether the effects of gamification on outcomes of interest varied according to the number of training sessions performed. There was no significant relationship between the number of training sessions and effect sizes for motivation/engagement outcomes (point estimate of slope=0.0002; 95% CI 0.12 to 0.12; *P*=.99; Multimedia Appendix 3) and cognitive process outcomes (point estimate of slope=0.01; 95% CI 0.07 to 0.09; *P*=.78; Multimedia Appendix 3). Of note is the low variability in the number of training sessions.

Quality of Included Studies

The risk of bias judgments for each included study is summarized in Table 4. Overall, the quality of the included RCTs was not optimal. The most common risks of bias were inadequate blinding of participants, personnel, and outcome assessors. The results did, however, reveal a low risk of bias for random sequence generation and selective reporting. It is noteworthy that for handling incomplete outcome data, all studies (with the exception of that of Mohammed et al [53]) did not provide the information necessary for assessing whether the criteria were met.

Table 4. Risk of bias assessment of the included studies.

Study[a]	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)
Boendermaker et al [47] study 1	Low	Low	High	High	Unclear	Low
Boendermaker et al [54]	Low	Low	High	High	Unclear	Low
Boendermaker et al [55]	Low	Low	High	High	Unclear	Low
Choi and Medalia [56]	Low	High	High	High	Unclear	Low
Dorrenbacher et al [57]	High	High	High	High	Unclear	Unclear
Katz et al [8]	High	High	Unclear	Unclear	Unclear	Unclear
Mohammed et al [53]	Low	High	Low	Low	Low	Low
Ninaus et al [18]	Low	Unclear	High	Low	Unclear	High
Prins et al [25]	Unclear	Unclear	High	Unclear	Unclear	Unclear

[a]Risk of bias assessment using the Cochrane risk of bias tool.

Discussion

Principal Findings

The aims of this study were to provide a state-of-the-art review of the gamification of cognitive training tasks and to evaluate its effectiveness. The systematic review identified 49 independent studies, with the majority published after 2010 (n=46/49, 94%) and conducted in European countries (n=32/49, 65%). The most comparable study is a systematic review by Lumsden et al [10] of gamified cognitive assessment and training tasks, which identified 33 studies published between 2007 and 2015. This indicates that the gamification of cognitive training is a rapidly evolving area of research. However, little is known about the effectiveness of such interventions. Only 9 (18%) of the 49 studies allowed rigorous testing of intervention effectiveness and were included in the meta-analysis, showing that the field is still developing and lacks well-controlled empirical studies. Overall, results from the meta-analysis of RCTs showed that gamified training tasks were more motivating/engaging and more demanding/difficult than non- or less-gamified tasks. No other significant effects were found, and no moderator variables were identified. However, the findings were limited by the small number of studies. The following sections discuss the key results obtained from the systematic review and meta-analysis and provide directions for future research.

Effectiveness of Gamification Applied to Cognitive Training

Gamification is commonly framed as a technique for increasing motivation and engagement in a given task [28]. The results from the meta-analysis suggest significant and moderate positive effects of gamification on motivation/engagement outcomes (N=8 studies; Hedges g=0.72; *P*=.002). This provides the first evidence that the application of gamification in cognitive training can positively influence motivation/engagement and confirms the findings of an earlier review by Lumsden et al [10], which also found evidence for the potential of gamification to enhance motivation and engagement in cognitive tasks. Our analysis, however, also indicates that gamified tasks are perceived as significantly more demanding/difficult. This is relatively unsurprising given that implementing game features adds a layer of complexity to a task [58]. However, the result is based on the synthesis of only 3 studies, and it remains unclear whether this affects training domain outcomes. Therefore, further research is warranted in this area.

Interestingly, we found no significant effect of gamification on cognitive outcomes (N=9 studies; Hedges g=0.27; *P*=.14). Thus, overall, gamification does not seem to have enhancing or deleterious effects on cognitive performance. Previous systematic reviews in the health [17] and cognitive [10] fields have reported that the evidence for the impact of gamification on cognitive outcomes is mixed. Some critical views on gamification interpret such findings as an illustration that it does not work [59], whereas others argue that these mixed results are due to confounders such as poor design [60,61] or failure to apply gamification in an appropriately considered or meaningful way [15]. In fact, the implementation of gamification has been cataloged within the group of complex interventions [60]. Complex interventions are activities that comprise multiple interacting components (eg, intensity, setting) that, when applied to the target population, result in variable outcomes [62]. Therefore, it is recommended that researchers developing and evaluating gamified health-related interventions consider using a framework, such as the Medical Research Council framework for complex interventions [62,63], to ensure that the implementation of game features has been well-thought through, and therefore, be more likely to produce the desired outcomes.

For both motivation/engagement and cognitive process outcomes, we found considerable heterogeneity in the results, which we were unable to explain by a systematic consideration of the age group, target population, number of game elements, and number of training sessions. Owing to insufficient studies, we were unable to examine the types of game elements or cognitive domains as potential moderators of outcomes. The results of these moderator analyses are further discussed in the following sections.

Audiences and Age Groups Investigated

In the systematic review, studies generally had small sample sizes, ranging widely from 3 [35,36] to 794 [37] participants, with nearly half (n=22/49, 45%) of the studies having fewer than 50 participants in total. In line with previous reviews [10,17], we found that gamification was applied to a large range of age groups, from children ([64], study 2) through adolescents [55] to adults [65]. Adults were the main target audience (n=38/49, 78%); however, we also found that more research findings have started to attach importance to the use of gamification for children and adolescents since 2011. With regard to the type of population investigated, most (n=33/49, 67%) of the interventions were delivered to high-risk groups, such as older adults [41] and individuals with specific health issues such as ADHD [66] or moderate levels of trait anxiety [67]. The remaining studies targeted low-risk samples such as university students [18] and school children [38].

In the meta-analysis, the findings that neither age nor population type had any moderating effects on motivation/engagement and cognitive process outcomes indicate that gamification may be equally applicable to a broad range of individuals. However, this does not imply that users should be treated as a monolithic group nor that researchers should adopt a *one-size-fits-all* design approach (ie, everyone interacts with the same gamification elements). Users differ in both behavior and motivation; therefore, gamified experiences should be tailored to their characteristics and contexts to maximize gamification effects [68-71]. For example, 2 training tasks may be gamified through the incorporation of a theme; however, the theme should be adapted to suit the target audience. In the study of Boendermaker et al [55], it is clear that the lighthearted "boy-meets-girl" love story theme is suitable for adolescents, but it may not necessarily be appropriate for children and adults. To date, the study of *personalized* gamification, while emerging, is still largely underexplored, and more research is needed to determine how best to satisfy individual and group differences in an effective way [72-75].

Cognitive Domains Targeted

In the systematic review, all training tasks were designed to improve various aspects of cognition. About two-thirds (n=31/49, 63%) of the studies conducted single-domain cognitive training, with the most targeted domain being attention (n=11/49, 22%), closely followed by working memory (n=9/49, 18%). This may be due to the fact that attentional processes and working memory are critical aspects of our cognitive capacities [76] that can be improved and modified with training [77,78]. The remaining assessed studies trained multiple cognitive domains simultaneously, making it difficult to isolate the impact of gamification on a specific domain. Overall, these findings are largely consistent with the results of Lumsden et al [10] who found gamified cognitive training interventions focused on one or two domains exclusively, mostly working memory.

Intervention Characteristics

In the systematic review, the type of technology used for delivering training tasks varied across studies, with computers (ie, desktop, laptop, and notebook) being the most widely used device (n=26/49, 53%), followed by tablets (n=13/49, 27%). Remarkably, only 2 studies employed smartphones [45,79], showing that the application of this technology to gamified cognitive training is still in its infancy. Unsurprisingly, therefore, training was mostly (n=19/49, 39%) delivered in a more controlled environment like the laboratory, hospital, or school. Of note, there is a lack of studies directly comparing the effects of the same intervention across different settings or delivery methods. To better understand the specific conditions for effective dissemination of gamified cognitive training interventions, future studies should investigate whether delivering training remotely and via mobile platforms (eg, smartphones, tablets) is acceptable and feasible. For example, it would be interesting to compare training outcomes (ie, cognitive and clinical) and adherence rates of a remote smartphone-based gamified intervention with an in-person desktop-based intervention in a randomized trial.

Overall, the number of training sessions ranged from only one session to a maximum of 59 sessions, for periods ranging from 5 min ([47], study 1) [54] up to 80 min [35] per training session. Previous research on cognitive training has shown a dose-response relationship between the number of training sessions and cognitive gains [80]. However, in the current meta-analysis, the number of sessions did not predict cognitive effects. It is possible that, in the context of gamification, more training sessions are needed, as there might be an *adaptation* time required to become used to the gamified environment. Meta-regression analyses also revealed no significant effect of the number of training sessions on effect sizes for motivation/engagement outcomes. The nonsignificant result of this moderator for cognitive process and motivation/engagement outcomes could be explained by the limited variability in the number of training sessions (Multimedia Appendix 3), making it difficult to detect a dose-response relationship.

Implementation of Gamification

In the systematic review, a broad range of game features have been integrated into training tasks, with the most frequent provision of elements involving achievement and progression-oriented features (see Table 1 for the full list). Specifically, a notable 82% (n=40/49) of all included studies incorporated rewards, such as points, badges, and stars, among others. This finding is consistent with previous systematic reviews [10,17], indicating that these features continue to dominate the landscape of gamification. Perhaps this is because rewards can easily be added as an additional layer to an existing paradigm without changing the original structure [81]. However, their use is not without critique, with researchers arguing that reward-based systems are motivating extrinsically, yet not intrinsically [13,15] with associated concerns regarding the longevity of any effects. Feedback loops were also frequently (n=39/49, 80%) employed as a means to amplify or reduce an action, and were delivered through various modalities, such as auditory (eg, high-pitched “Huh?” [82]), visual (eg, growing neuron animation [18]), and tactile (eg, vibration on the smartphone [79]). This observation is unsurprising because feedback is a key technique for behavior change [83] and can facilitate intrinsic motivation [84].

The second most frequently implemented category of game elements involves immersion-oriented features, such as story/theme, sound effects, and avatars (Table 1). In particular, the use of a story/theme was found to be high, with more than three-fourth of studies (n=38/49) using this technique to enhance the appeal of cognitive tasks. In some studies, it was evident that the theme was carefully selected to fit the target audience. For example, Dassen et al [85] presented their working memory tasks in a restaurant setting to help overweight adults improve self-regulation and increase weight loss. The high frequency of immersion-related features is encouraging because, from an SDT perspective [12,13], such elements can potentially satisfy the need for autonomy and drive intrinsic motivation [86]. On the other hand, social-oriented features were underutilized, with only a few studies favoring competition (n=3/49, 6%) and social interaction (n=2/49, 4%). Similar findings have been reported in a recent review investigating game-based interventions for neuropsychological assessment, training, and rehabilitation [87]. This may be due to implementation complexity or the relative recency of the field.

It is clear from our review that studies are driven by the presumption that gamification in cognitive training consists of embedding a combination of game elements within the training tasks. Indeed, no study has reported on the effect of a single element alone and the mode count of game elements was 5 (range 2-9). This figure is much higher than from a previous review of web-based mental health interventions (mode 1; range 1-3) [88], but is in line with a recent review of mental health and well-being apps (mode 5; range 1-11) [89], highlighting that the nature of gamification in the context of health is changing, becoming increasingly complex and diverse.

In the meta-analysis, we attempted to include the type of game elements used as a moderator; however, there were not enough studies in the subgroups for a conclusive evaluation (Tables 2 and 3). Meta-regression analyses revealed no significant effect of the number of game elements on effect sizes for motivation/engagement and cognitive process outcomes. This observation raises the question of whether there is a minimum number of gamification features that need to be implemented to achieve the threshold for motivation/engagement (without significantly impacting cognitive processes). A related question is when a gamified task should be considered to have become a game—the boundaries between gamification and serious games remain blurry and highly subjective [10,17,89].

With regard to theories of motivation, only around one-third of studies (n=17/49, 35%) in our systematic review mentioned using theoretical foundations to guide the development of their interventions. SDT [12,13], flow theory [50], and the framework by Kiili and Perttula [51] were among the most prominent choices of frameworks (n=7/49, 14%) used in gamified cognitive training. Of note, several studies (n=8/49, 16%) drew upon multiple theories and principles to design their training tasks. This lack of theoretical underpinning has been reported previously [14,17], calling for more theory-driven research on gamification in the field of health.

Limitations

There are limitations to this study. First, the results of the systematic review and meta-analysis are limited by the relatively low methodological quality of retrieved studies, mostly due to inadequate or unclear blinding of participants and personnel. Previous reviews [17] have provided similar findings, reflecting the relative infancy of the gamification field in the health context. There was also a large degree of heterogeneity between studies including study design, target population, and outcome measures, which created challenges in synthesizing the literature and may have contributed to the heterogeneity detected in the meta-analyses. Of particular concern was the widespread use of self-developed, unvalidated tools to assess motivation and engagement outcomes. Thus, to facilitate interpretation of results and to advance gamification research, there is a need for experts and researchers to develop valid, reliable, and sensitive measures of motivation and engagement that are grounded in motivational theory and are applicable to the particular research and clinical context. Another limitation of the review was the modest number of studies (n=9/49, 18%) that we could include in the meta-analyses. In addition, sample sizes were rather small, especially for several process-related outcomes and subgroup analyses, limiting the generalizability of our results. Owing to the limited number of RCTs, we were also unable to test for the cognitive domain and type of game elements as potential moderators of gamification effects. The size and direction of the effects of gamification on cognitive outcomes may differ by domain; the examination of which requires more RCTs to be conducted in the domains of attention and inhibition. It is possible that the absence of evidence of any effect of gamification on cognitive outcomes overall might reflect the relatively small number of heterogeneous studies and not that gamification cannot be applied in a manner that will improve cognitive outcomes in particular domains. With the addition of further studies and associated variability, future meta-analyses may want to assess the influence of such moderators and potential sources of bias (eg, funding source) on review results and conclusions. Available study findings also did not address the long-term or sustained effects of gamified interventions due to limited follow-up assessments of outcomes (n=1/9, 11%) [54]. Finally, as mentioned previously, game elements were not always well-described in studies and investigated only in combination, making it impossible to establish whether individual elements had measurable effects.

Recommendations for Future Research

Exploration of Game Elements

Our review revealed that a wide array of game elements is being used; however, certain features (eg, rewards) continue to receive more attention than others (eg, social interactions). The further development of gamified tasks may benefit from drawing from the entire repertoire of game elements and tailoring gamification according to the users' individual characteristics and contexts. Furthermore, consistent with previous reviews [17], there is a lack of research isolating the impact of single game elements. However, given that in real-world settings, gamification elements are not typically deployed in isolation, it may be most beneficial for future research to explore both the impact of individual elements as well as groups of elements.

Reporting of Gamification

Much progress can be made in improving the reporting of gamification elements, with only a few studies in our review explicitly describing how each element is operationalized in the interventions. As such, we propose the creation of reporting guidelines that outline a full description of the game elements used (via text and picture) in research, how they are being implemented, and what they aim to target. To facilitate understanding of which game elements are used and how they are operationalized, researchers should enable the scientific community to consult a (sample) version of the gamified task. This will improve the quality of gamified health-related publications and facilitate more informative systematic reviews.

Reviewing the Quality of Gamification

Relatedly, not only the reporting but also the evaluation of the quality of the operationalization of gamification is important. Therefore, we call for gamification developers to create guidelines for evaluating the quality of the intervention. Given that it is unlikely that such an evaluation can be made based on the information provided in the paper, gamified tasks should be more readily accessible to other researchers. Future research could serve the field well by *archiving* gamified interventions and by implementing effective data management strategies that use the Findable, Accessible, Interoperable, and Reusable principles [90].

Perceptions Toward Gamification

Surprisingly, almost no studies have investigated perceptions of different populations regarding the use of gamification. Indeed, while several studies have indicated that participants experience traditional cognitive training tasks as monotonous and boring [4,5], more research is needed to evaluate participants' perceptions and readiness for gamification before determining whether it should be implemented. One way of achieving this is by using a combination of quantitative and qualitative methods to determine the target population's perceptions and experiences of gamification as well as its receptiveness to using gamification in such tasks. This approach could also improve tailoring gamification to specific target groups. For example, adolescents (generation Z) and young adults (millennials) may be more responsive to gamification experience given their familiarity with a broad range of digital technologies.

Theory-Driven Research

Most reviewed studies lack a theoretical underpinning for the choice and design of their interventions. This issue is not new and has been reported previously [14,17], calling for more theory-driven research on gamification in the field of cognition and health more generally. Future research should develop and establish more formal working models of gamification that help understand *why* and *how* particular game elements work (or not work), when they work best, and the kind of effects they expect to have on psychological or behavioral outcomes.

Study Design and Reporting

Despite growing interest in applying gamification to cognitive training, there is still a lack of long-term follow-up and well-controlled studies in this field. More rigorously designed RCTs comparing gamified and nongamified versions of the same intervention, with adequately powered sample sizes and longer-term follow-up are needed. We also urge researchers to preregister research protocols and conform to the recommendations of the Consolidated Standards of Reporting Trials guidelines [91].

Acknowledgments

The authors thank all researchers who kindly answered our inquiries and made the meta-analysis possible.

Authors' Contributions

All authors were involved in the design and development of the study protocol, which was undertaken as part of a PhD for JV under the supervision of the other authors. JV conducted a literature search. JV and DV coded the included studies, in consultation with the other authors. JV conducted the statistical analysis and wrote the first draft of the paper. All authors contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategies.
[PDF File (Adobe PDF File), 64 KB - games_v8i3e18644_app1.pdf]

Multimedia Appendix 2
Systematic review tables.
[PDF File (Adobe PDF File), 284 KB - games_v8i3e18644_app2.pdf]

Multimedia Appendix 3
Meta-analysis tables and figures.
[PDF File (Adobe PDF File), 327 KB - games_v8i3e18644_app3.pdf]

References

1. Harvey PD, McGurk SR, Mahncke H, Wykes T. Controversies in computerized cognitive training. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Nov;3(11):907-915 [FREE Full text] [doi: 10.1016/j.bpsc.2018.06.008] [Medline: 30197048]
2. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
3. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
4. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
5. Kuckertz JM, Schofield CA, Clerkin EM, Primack J, Boettcher H, Weisberg RB, et al. Attentional bias modification for social anxiety disorder: what do patients think and why does it matter? Behav Cogn Psychother 2019 Jan;47(1):16-38 [FREE Full text] [doi: 10.1017/S1352465818000231] [Medline: 29729676]
6. Jaeggi SM, Buschkuehl M, Shah P, Jonides J. The role of individual differences in cognitive training and transfer. Mem Cognit 2014 Apr;42(3):464-480. [doi: 10.3758/s13421-013-0364-z] [Medline: 24081919]
7. Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, et al. No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. J Exp Psychol Gen 2013 May;142(2):359-379. [doi: 10.1037/a0029082] [Medline: 22708717]
8. Katz B, Jaeggi S, Buschkuehl M, Stegman A, Shah P. Differential effect of motivational features on training improvements in school-based cognitive training. Front Hum Neurosci 2014;8:242 [FREE Full text] [doi: 10.3389/fnhum.2014.00242] [Medline: 24795603]
9. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
10. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
11. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; October 6-8, 2011; Tampere, Finland p. 28-30. [doi: 10.1145/2181037.2181040]
12. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, USA: Springer Science & Business Media; 1985.
13. Ryan RM, Deci E. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
14. Seaborn K, Fels DI. Gamification in theory and action: a survey. Inte JHum-Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
15. Nicholson S. A RECIPE for meaningful gamification. In: Reiners T, Wood LC, editors. Gamification in Education and Business. Switzerland: Springer International Publishing; 2015:1-20.
16. Muntean CI. Raising Engagement in E-Learning Through Gamification. In: Proceedings of the 6th International Conference on Virtual Learning. 2011 Presented at: ICVL'11; October 28-29, 2011; Romania, Europe URL: http://icvl.eu/2011/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper42.pdf [doi: 10.15837/ijccc.2007.1.2334]
17. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]

18. Ninaus M, Pereira G, Stefitz R, Prada R, Paiva A, Neuper C, et al. Game elements improve performance in a working memory training task. Int J Serious Games 2015 Feb 10;2(1):3-16. [doi: 10.17083/ijsg.v2i1.60]
19. Lee T, Goh SJ, Quek SY, Phillips R, Guan C, Cheung YB, et al. A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy. PLoS One 2013;8(11):e79419 [FREE Full text] [doi: 10.1371/journal.pone.0079419] [Medline: 24260218]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. Johnson D, Horton E, Mulcahy R, Foth M. Gamification and serious games within the domain of domestic energy consumption: a systematic review. Renew Sust Energ Rev 2017 Jun;73:249-264. [doi: 10.1016/j.rser.2017.01.134]
22. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
23. Higgins J, Li T, Deeks J. Choosing effect measures and computing estimates of effect. In: Higgins J, Thomas J, Chandler J, Cumpston M, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. Second Edition. Chichester, UK: John Wiley & Sons Ltd; 2019.
24. van Ryckeghem DM, van Damme S, Eccleston C, Crombez G. The efficacy of attentional distraction and sensory monitoring in chronic pain patients: a meta-analysis. Clin Psychol Rev 2018 Feb;59:16-29. [doi: 10.1016/j.cpr.2017.10.008] [Medline: 29126746]
25. Prins PJ, Dovis S, Ponsioen A, ten Brink E, van der Oord S. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? Cyberpsychol Behav Soc Netw 2011 Mar;14(3):115-122. [doi: 10.1089/cyber.2009.0206] [Medline: 20649448]
26. Higgins J, Altman D, Sterne J. Assessing risk of bias in included studies. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. West Sussex, England: John Wiley & Son Ltd; 2011.
27. Koivisto J, Hamari J. Demographic differences in perceived benefits from gamification. Comput Hum Behav 2014 Jun;35:179-188. [doi: 10.1016/j.chb.2014.03.007]
28. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inf Manag 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
29. Hedges L, Olkin I. Statistical Methods for Meta-Analysis. San Diego, CA: Academic Press; 1985.
30. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to Meta-Analysis. Chichester, United Kingdom: John Wiley & Sons; 2009.
31. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002 Jun 15;21(11):1559-1573. [doi: 10.1002/sim.1187] [Medline: 12111920]
32. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
33. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
34. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
35. Brezovszky B, Lehtinen E, McMullen J, Rodriguez G, Veermans K. Training Flexible and Adaptive Arithmetic Problem Solving Skills Through Exploration With Numbers: the Development of Numbernavigation Game. In: Proceedings of the 7th European Conference on Games Based Learning. 2013 Presented at: ECGBL'13; October 3-4, 2013; Porto, Portugal URL: https://tinyurl.com/yyv4kxk4
36. Connor B, Shaw C. Case study series using brain-training games to treat attention and memory following brain injury. J Pain Manag 2016;9(3):217-226 [FREE Full text]
37. Double KS, Birney DP. The effects of personality and metacognitive beliefs on cognitive training adherence and performance. Pers Individ Differ 2016 Nov;102:7-12. [doi: 10.1016/j.paid.2016.04.101]
38. Jaeggi SM, Buschkuehl M, Jonides J, Shah P. Short- and long-term benefits of cognitive training. Proc Natl Acad Sci U S A 2011 Jun 21;108(25):10081-10086 [FREE Full text] [doi: 10.1073/pnas.1103228108] [Medline: 21670271]
39. Nagle A, Novak D, Wolf P, Riener R. Increased enjoyment using a tablet-based serious game with regularly changing visual elements: a pilot study. Gerontechnology 2015;14(1):32-44 [FREE Full text] [doi: 10.4017/gt.2015.14.1.001.00]
40. De Vries M, Prins PJ, Schmand BA, Geurts HM. Working memory and cognitive flexibility-training for children with an autism spectrum disorder: a randomized controlled trial. J Child Psychol Psychiatry 2015 May;56(5):566-576. [doi: 10.1111/jcpp.12324] [Medline: 25256627]
41. Hiraoka T, Wang T, Kawakami H. Cognitive function training system using game-based design for elderly drivers. IFAC-PapersOnLine 2016;49(19):579-584 [FREE Full text] [doi: 10.1016/j.ifacol.2016.10.613]
42. Dennis-Tiwary TA, Denefrio S, Gelber S. Salutary effects of an attention bias modification mobile application on biobehavioral measures of stress and anxiety during pregnancy. Biol Psychol 2017 Jul;127:148-156 [FREE Full text] [doi: 10.1016/j.biopsycho.2017.05.003] [Medline: 28478138]

43. Boletsis C, McCallum S. Augmented reality cubes for cognitive gaming: preliminary usability and game experience testing. Int J Serious Games 2016 Mar 1;3(1):3-18. [doi: 10.17083/ijsg.v3i1.106]
44. Enock P. Making an IMPACT: Designing and Testing a Novel Attentional Training Game to Reduce Social Anxiety. Harvard University. 2015. URL: https://dash.harvard.edu/handle/1/17467292 [accessed 2018-07-13]
45. Garolera M, Berga N, Quintana M, Chico G, Cerulla N, López M, et al. Active-U: Playing to Stimulate your Brain. In: Proceedings of the 2nd International Workshop on Gamification in Health. 2015 Nov Presented at: gHealth'15; 10-13 November 2015; Barcelona, Spain p. 2-7 URL: http://ceur-ws.org/Vol-1547/paper1.pdf
46. Abellanoza C. Rewind-Remind: Investigating How Gamification of Memory Tasks Can Evaluate Associative Memory Performance in Healthy, Older Adults. The University of Texas at Arlington. 2017. URL: https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/26785/ABELLANOZA-DISSERTATION-2017.pdf?sequence=1 [accessed 2018-07-16]
47. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
48. Trapp W, Landgrebe M, Hoesl K, Lautenbacher S, Gallhofer B, Günther W, et al. Cognitive remediation improves cognition and good cognitive performance increases time to relapse--results of a 5 year catamnestic study in schizophrenia patients. BMC Psychiatry 2013 Jul 9;13:184 [FREE Full text] [doi: 10.1186/1471-244X-13-184] [Medline: 23837673]
49. Birk MV, Mandryk RL, Atkins C. The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY'16; October 26-31, 2016; Austin, Texas, USA. [doi: 10.1145/2967934.2968091]
50. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York, USA: Harper & Row; 1990.
51. Kiili K, Perttula A. A design framework for educational exergames. In: Freitas S, Ott M, Popescu MM, Stanescu I, editors. New Pedagogical Approaches in Game Enhanced Learning: Curriculum Integration. Hershey, PA, USA: IGI Global; 2013:136-158.
52. Gee JP. Learning by design: good video games as learning machines. E-Learn Digit Media 2016 Jul 31;2(1):5-16. [doi: 10.2304/elea.2005.2.1.5]
53. Mohammed S, Flores L, Deveau J, Hoffing RC, Phung C, Parlett CM, et al. The benefits and challenges of implementing motivational features to boost cognitive training outcome. J Cogn Enhanc 2017 Dec;1(4):491-507 [FREE Full text] [doi: 10.1007/s41465-017-0047-y] [Medline: 30221244]
54. Boendermaker WJ, Maceiras SS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 6;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
55. Boendermaker WJ, Veltkamp RC, Peeters M. Training behavioral control in adolescents using a serious game. Games Health J 2017 Dec;6(6):351-357. [doi: 10.1089/g4h.2017.0071] [Medline: 28926286]
56. Choi J, Medalia A. Intrinsic motivation and learning in a schizophrenia spectrum sample. Schizophr Res 2010 May;118(1-3):12-19 [FREE Full text] [doi: 10.1016/j.schres.2009.08.001] [Medline: 19716270]
57. Dörrenbächer S, Müller PM, Tröger J, Kray J. Dissociable effects of game elements on motivation and cognition in a task-switching training in middle childhood. Front Psychol 2014;5:1275 [FREE Full text] [doi: 10.3389/fpsyg.2014.01275] [Medline: 25431564]
58. Lopez CE, Tucker CS. A quantitative method for evaluating the complexity of implementing and performing game features in physically-interactive gamified applications. Comput Hum Behav 2017 Jun;71:42-58. [doi: 10.1016/j.chb.2017.01.036]
59. Bogost I. Why gamification is bullshit. In: Walz SP, Deterding S, editors. The Gameful World: Approaches, Issues, Applications. Cambridge, MA: MIT Press; 2015:65-80.
60. Rojas D, Kapralos B, Dubrowski A. The Missing Piece in the Gamification Puzzle. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications. 2013 Presented at: Gamification'13; October 2-4, 2013; Stratford, Canada. [doi: 10.1145/2583008.2583033]
61. Wen C, Hsien H, Huang H. A Survey of Gamification for Healthcare System. In: Proceedings of the International Conference on Orange Technologies. 2015 Presented at: ICOT'15; December 19-22, 2015; Hong Kong, China. [doi: 10.1109/icot.2015.7498491]
62. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new medical research council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
63. Campbell M, Fitzpatrick R, Haines A, Kinmonth A, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
64. Kiili K, Ninaus M, Koskela M, Tuomi M, Lindstedt A. Developing Games for Health Impact: Case Brains vs Zombies. In: Proceedings of the 7th European Conference on Games Based Learning. 2013 Presented at: ECGBL'13; October 3-4, 2013; Porto, Portugal URL: https://www.researchgate.net/publication/272480884_Developing_games_for_health_impact_Case_brains_vs_zombies
65. Dennis-Tiwary TA, Egan LJ, Babkirk S, Denefrio S. For whom the bell tolls: neurocognitive individual differences in the acute stress-reduction effects of an attention bias modification game for anxiety. Behav Res Ther 2016 Feb;77:105-117 [FREE Full text] [doi: 10.1016/j.brat.2015.12.008] [Medline: 26745621]

66. van der Oord S, Ponsioen AJ, Geurts HM, ten Brink EL, Prins PJ. A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: outcome on parent- and teacher-rated executive functioning and ADHD behavior. J Atten Disord 2014 Nov;18(8):699-712. [doi: 10.1177/1087054712453167] [Medline: 22879577]
67. Notebaert L, Clarke PJ, Grafton B, MacLeod C. Validation of a novel attentional bias modification task: the future may be in the cards. Behav Res Ther 2015 Feb;65:93-100. [doi: 10.1016/j.brat.2014.12.007] [Medline: 25594940]
68. Kappen D, Mirza-Babaei P, Nacke L. Gamification through the Application of Motivational Affordances for Physical Activity Technology. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY'17; October 15-18, 2017; Amsterdam, The Netherlands. [doi: 10.1145/3116595.3116604]
69. Codish D, Ravid G. Personality Based Gamification- Educational Gamification for Extroverts and Introverts. In: Proceedings of the 9th CHAIS Conference for the Study of Innovation and Learning Technologies: Learning in the Technological Era. 2014 Presented at: CHAIS'14; December 11, 2014; Raanana, Israel p. 36-44 URL: https://www.openu.ac.il/innovation/chais2014/download/E2-2.pdf
70. Orji R, Mandryk R, Vassileva J, Gerling K. Tailoring Persuasive Health Games to Gamer Type. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France. [doi: 10.1145/2470654.2481341]
71. Böckle M, Novak J, Bick M. Towards Adaptive Gamification: A Synthesis of Current Developments. In: Proceedings of the 25th European Conference on Information Systems. 2017 Presented at: ECIS'17; June 5-10, 2017; Guimarães, Portugal URL: https://www.researchgate.net/publication/317687382_TOWARDS_ADAPTIVE_GAMIFICATION_A_SYNTHESIS_OF_CURRENT_DEVELOPMENTS
72. Knutas A, van Roy R, Hynninen T, Granato M, Kasurinen J, Ikonen J. A process for designing algorithm-based personalized gamification. Multimed Tools Appl 2018 Dec 5;78(10):13593-13612. [doi: 10.1007/s11042-018-6913-5]
73. Oliveira W, Bittencourt I. Research advances on tailored gamification. In: Tailored Gamification to Educational Technologies. Singapore: Springer Nature; Oct 20, 2019:21-31.
74. Tondello G, Wehbe R, Diamond L, Busch M, Marczewski A, Nacke L. The Gamification User Types Hexad Scale. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY'16; October 16-19, 2016; Austin, USA. [doi: 10.1145/2967934.2968082]
75. Hallifax S, Serna A, Marty J, Lavoué G, Lavoué E. Factors to Consider for Tailored Gamification. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2019 Presented at: CHI PLAY'19; September 2-6, 2019; Paphos, Cyprus. [doi: 10.1145/3311350.3347167]
76. Fougnie D. The relationship between attention and working memory. In: Johansen NB, editor. New Research on Short-Term Memory. New York, USA: Nova Science Publishers; 2008:1-45.
77. Zhang H, Huntley J, Bhome R, Holmes B, Cahill J, Gould RL, et al. Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: a systematic review and meta-analysis. BMJ Open 2019 Aug 18;9(8):e027062 [FREE Full text] [doi: 10.1136/bmjopen-2018-027062] [Medline: 31427316]
78. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 1;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
79. Lukas CA, Berking M. Reducing procrastination using a smartphone-based treatment program: a randomized controlled pilot study. Internet Interv 2018 Jun;12:83-90. [doi: 10.1016/j.invent.2017.07.002] [Medline: 30135772]
80. Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proc Natl Acad Sci U S A 2008 May 13;105(19):6829-6833 [FREE Full text] [doi: 10.1073/pnas.0801268105] [Medline: 18443283]
81. Boendermaker WJ, Prins PJ, Wiers RW. Cognitive bias modification for adolescents with substance use problems--can serious games help? J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):13-20. [doi: 10.1016/j.jbtep.2015.03.008] [Medline: 25843611]
82. Dennis TA, O'Toole L. Mental health on the go: effects of a gamified attention bias modification mobile application in trait anxious adults. Clin Psychol Sci 2014 Sep 1;2(5):576-590. [doi: 10.1177/2167702614522228] [Medline: 26029490]
83. 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. [doi: 10.2196/games.3139] [Medline: 25658754]
84. Hicks K, Gerling K, Richardson G, Pike T, Burman O, Dickinson P. Understanding the Effects of Gamification and Juiciness on Players. In: Proceedings of the IEEE Conference on Games. 2019 Presented at: CoG'19; August 20-23, 2019; London, United Kingdom. [doi: 10.1109/cig.2019.8848105]
85. Dassen FC, Houben K, van Breukelen GJ, Jansen A. Gamified working memory training in overweight individuals reduces food intake but not body weight. Appetite 2018 May 1;124:89-98. [doi: 10.1016/j.appet.2017.05.009] [Medline: 28479405]
86. Xi N, Hamari J. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. Int J Inf Manag 2019 Jun;46:210-221. [doi: 10.1016/j.ijinfomgt.2018.12.002]
87. Ferreira-Brito F, Fialho M, Virgolino A, Neves I, Miranda AC, Sousa-Santos N, et al. Game-based interventions for neuropsychological assessment, training and rehabilitation: which game-elements to use? A systematic review. J Biomed Inform 2019 Oct;98:103287. [doi: 10.1016/j.jbi.2019.103287] [Medline: 31518700]

88. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
89. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
90. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
91. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CMA: Comprehensive Meta-Analysis
RCT: randomized controlled trial
SDT: self-determination theory

Edited by G Eysenbach; submitted 10.03.20; peer-reviewed by A Peckham, M Brown, M Behzadifar; comments to author 16.04.20; revised version received 21.04.20; accepted 22.04.20; published 10.08.20.

Please cite as:
Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML
The Effects of Gamification on Computerized Cognitive Training: Systematic Review and Meta-Analysis
JMIR Serious Games 2020;8(3):e18644
URL: https://games.jmir.org/2020/3/e18644
doi:10.2196/18644
PMID:32773374

Original Paper

Effective Gamification of the Stop-Signal Task: Two Controlled Laboratory Experiments

Maximilian Achim Friehs[1], PhD; Martin Dechant[2], MSc; Sarah Vedress[2], BA; Christian Frings[1], PhD; Regan Lee Mandryk[2], PhD

[1]Department of Cognitive Psychology and Methodology, Trier University, Trier, Germany

[2]Human-Computer-Interaction Lab, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:
Maximilian Achim Friehs, PhD
Department of Cognitive Psychology and Methodology
Trier University
Universitätsring 15
Trier, 54292
Germany
Phone: 49 651201 ext 3250
Email: friehs@uni-trier.de

Abstract

Background: A lack of ability to inhibit prepotent responses, or more generally a lack of impulse control, is associated with several disorders such as attention-deficit/hyperactivity disorder and schizophrenia as well as general damage to the prefrontal cortex. A stop-signal task (SST) is a reliable and established measure of response inhibition. However, using the SST as an objective assessment in diagnostic or research-focused settings places significant stress on participants as the task itself requires concentration and cognitive effort and is not particularly engaging. This can lead to decreased motivation to follow task instructions and poor data quality, which can affect assessment efficacy and might increase drop-out rates. Gamification—the application of game-based elements in nongame settings—has shown to improve engaged attention to a cognitive task, thus increasing participant motivation and data quality.

Objective: This study aims to design a gamified SST that improves participants' engagement and validate this gamified SST against a standard SST.

Methods: We described the design of our gamified SST and reported on 2 separate studies that aim to validate the gamified SST relative to a standard SST. In study 1, a within-subject design was used to compare the performance of the SST and a stop-signal game (SSG). In study 2, we added eye tracking to the procedure to determine if overt attention was affected and aimed to replicate the findings from study 1 in a between-subjects design. Furthermore, in both studies, flow and motivational experiences were measured.

Results: In contrast, the behavioral performance was comparable between the tasks (P<.87; BF01=2.87), and the experience of flow and intrinsic motivation were rated higher in the SSG group, although this difference was not significant.

Conclusions: Overall, our findings provide evidence that the gamification of SST is possible and that the SSG is enjoyed more. Thus, when participant engagement is critical, we recommend using the SSG instead of the SST.

(JMIR Serious Games 2020;8(3):e17810) doi:10.2196/17810

KEYWORDS

video games; games, experimental; proof of concept study; cognition; psychology; motivation

Introduction

Background

Gamification is the process of applying game design elements (eg, scoring systems, graphical interface, narrative) to nongame environments (eg, cognitive tasks, work context) to increase task performance and engagement [1]. Gamification has been used in a variety of settings, such as in business [2] and education [3]. Serious games are also used in the context of health care education to support desirable behavior [1,4-7]. The use of games or game-like tasks makes it possible to enhance voluntary engagement and decrease participant drop-out rates [8,9]; in fact, a recent study showed that the experience of playing digital games as compared with standard cognitive tasks was perceived as less stressful [10]. A high dropout rate, especially in difficult-to-obtain samples, can lead to difficulties interpreting the results, for example, due to decreased statistical power [11-14]. Increased task engagement is especially important when cognitive tasks are used as a diagnostic tool because they rely upon the participant to perform the task to the best of their ability. Data obtained from individuals who lack the motivation to perform the task will not be representative of their ability, and this can lead to misinterpretations [15-17]. Although it seems that gamification is generally useful, it can also change task performance in an undesired direction [18]. For example, adding a simple scoring or reward system creates a motivational pull that can interact and interfere with the to-be-measured variable and change behavior [19-21]. However, a reward can even capture attention when it is counterproductive to the task performance, which might make simple reward elements, for example, not always suitable for all gamification purposes [22].

Cognitive Task Gamification

There have been efforts to gamify cognitive tasks for the purposes of training and assessment [23,24]. The interpretation of cognitive task data depends on the assumption that individuals are putting forth their best effort and are fully attentive to the task, but cognitive tasks are often repetitive and boring, so unfocused effort is a common problem [15,16]. An individual’s true ability will not be represented if they are not engaged and fully attentive, which can lead to inaccurate interpretations of cognitive task performance data [17]. To improve engagement with cognitive tasks, researchers have looked to games [25,26], with Aeberhard et al [27] noting that “leveraging gamification to repeatedly obtain behavioral samples paves the way for next-generation high-throughput psychometric toolset.”

However, caution must be taken when introducing game elements to cognitive tasks owing to the risk of muddying the measurement of the targeted cognitive process [28]. Cognitive tasks are very sensitive to manipulation—even basic tasks (eg, Stroop task, dot-probe task) are extensively studied to understand the effects of making small changes to the task paradigm [29]. Adding game-based elements to basic cognitive tasks could affect performance and experience in unintended ways [28]. Studies on how gamification of cognitive tasks affect behavior have shown mixed results. For example, adding points (a common gamification technique) to a task has been shown to increase engagement [28,30-32] and improve performance, such as by facilitating faster reaction times [24,28]. However, the inclusion of points has also shown to increase error rates in a dot-probe task [28]. Adding thematic elements and complex graphics has been shown to lead to decreased performance: for example, in a go-no-go task, the use of cowboy characters resulted in worse performance compared with a control (green and red objects) [31], and the use of zombie characters resulted in worse performance compared with a control (circles and squares) [33], likely because the stimuli were not as simple to discriminate. However, the inclusion of thematic elements and graphical stimuli have been shown to increase enjoyment [31] but also decrease enjoyment [28,30,33], relative to a control task.

As there is little agreement on how typical gamification approaches affect performance on, and engagement with, cognitive tasks [24,28], it is imperative that gamified cognitive tasks, intended for use in research, be validated against the basic version before use. Especially, in the context of cognitive psychology or clinical diagnostics, it is important to maintain internal validity [24,34].

Theoretical Underpinnings of Gamification

There are many theories that go beyond the mantra of “games are fun” as to why game design elements are so successful in shaping behavior. Although there is still an open debate regarding the understanding of what makes games enjoyable [35], two of the most prominent theories are the Flow Theory of Motivation [36] and the Self-Determination Theory (SDT) [37].

The Flow Theory states that there are some factors that facilitate flow experience. Specifically, the activity must have clear goals, there must be immediate and unambiguous feedback during task performance, and the perceived challenges of the activity must be balanced with the individual’s own skills [38-40]. A flow experience itself differs from individual to individual but is generally characterized by a high concentration on the task at hand, a loss of self-consciousness, a loss of sense of time, and deriving personal purpose from the task performance (ie, autotelic experience) [36,38-40]. In games and player experience research, flow is a key concept and has been proven, among other factors, to be important for player motivation and retention [41-46].

SDT is based upon 3 basic needs: the need for competence (ie, experiencing mastery over challenges); the need for autonomy (ie, doing something owing to an individual’s own volition); and the need for relatedness (ie, experiencing meaningful social relations) [47-49]. Importantly, games have been shown to be capable of addressing those needs and enhancing intrinsic motivation [37]. If one or ideally all 3 needs are satisfied, the motivation to engage in the task will increase [50-53]. SDT has been mirrored in the gamification classification system developed by Nicholson [54], in which he proposed 2 types of gamification: reward-based gamification and meaningful gamification. Although reward-based gamification aims to modify extrinsic motivation, meaningful gamification aims to increase intrinsic motivation. Thus, SDT can be used to explain the underlying components of intrinsic motivation, which has been shown to be an important predictor of task engagement [37,55].

In summary, flow theory and SDT are 2 promising theories that can explain an individual’s motivation for and experience while performing a task. Importantly, the 2 perspectives are not mutually exclusive but rather complement each other. Thus, gamification based on these theories can inform certain design guidelines for developing gamified versions of cognitive tasks. [54,56].

The Stop-Signal Task

One such cognitive task that is valuable to assess is the ability to inhibit an already initiated action. For example, a basketball player on defense might have to suppress his or her jumping response to avoid falling for the pump-fake of the offensive player or a person might have to stop crossing the street to avoid a speeding car. This type of response inhibition process can be measured using the stop-signal task (SST), which is an established measure of response inhibition and has been used in laboratories now for over 50 years [57,58]. The ability to inhibit a response is also modulated by inter- and intrapersonal differences in humans. For example, a reduction in inhibitory control and a general increase in impulsivity can be seen in people with attention-deficit/hyperactivity disorder (ADHD) [59,60] or patients with schizophrenia [61,62]. In addition, evidence suggests that training or certain types of sports [63,64], as well as noninvasive brain stimulation, can modulate an individual's ability to stop a response [65,66]. As response inhibition has been consistently associated with certain disorders, it has been proposed that response inhibition capabilities can be used as a form of objective diagnostic indicator, especially in ADHD but also in other disorders such as obsessive-compulsive disorder (OCD) [60,67-69]. Individuals affected by mental disorders, especially in the case of ADHD, may have problems focusing on the cognitive task, which makes it particularly important to develop a task that is more engaging to properly assess their cognitive functioning. However, consideration must be taken as gamified tasks have been shown to normalize the performance of individuals with ADHD, meaning that the gamified cognitive task no longer differentiates between people with and without ADHD [70].

The SST requires the participant to withhold their response on a random subset of trials during a choice reaction time task. The delay after which the stopping cue is presented (aptly termed stop-signal delay [SSD]) is fitted to the individual so that in approximately half of all stopping trials, the response inhibition will fail. In detail, when a participant successfully stops their response, the SSD is increased, making a successful stopping less likely on the subsequent trial (vice versa for unsuccessful stop-trials). Usually, participants are tested in a controlled, distraction-free environment, and the stimuli are presented on a monochromatic screen without any irrelevant or interfering elements. Although this leads to a very precise and clean measurement of an individual's response inhibition capabilities, it is not comparable with everyday situations in which the stopping of an already initiated response is required.

In other areas dealing with inhibition of information or responses, an effort has been made to transfer fundamental research principles to applied settings. For example, it was shown that the conflict resolution process as measured by classical cognitive psychological tasks such as the Stroop task or Eriksen flanker task [71,72] is conceptually similar and abides by the same rules as deceptive actions in sports [73-75]. Interestingly, recent studies provide evidence that even the underlying neural generators of these 2 conceptually analog tasks are similar [76,77]. However, as previously mentioned, caution must be taken when adding visual complexity to cognitive tasks due to the potential effects on performance.

In the case of the SST, previous work has shown that changing the stimuli from colored circles to colored fruits (along with an accompanying narrative) resulted in greater stop-signal reaction times (SSRTs; ie, worsened performance) relative to a version gamified with points, but no narrative [30]; however, enjoyment was also reduced in the thematic version relative to the points version, suggesting that engagement may not have been facilitated through the particular theme and stimuli chosen. Research on other tasks has suggested that a poorly implemented theme that offers little gameplay might be worse for engagement than including no theme at all [28].

This Study

To better understand everyday human behavior or, in the case of this paper, specifically stopping the behavior, gamification might be helpful to aid researchers in gathering large data sets over time. In this case, a gamified version of the SST would allow researchers to enhance the ecological validity of the inhibition measurement by presenting the task in a visually complex environment, while also keeping participants motivated to perform well. This ties into the proposition that modern technology can be used to enhance mundane and experimental realism while keeping experimental control high and potentially even increase the effect of experimental manipulation [78,79]. A gamified SST can mirror a more natural setting and therefore elicit more natural responses without sacrificing experimental control. Thus, it is important to choose a task design that not only reliably taps into the targeted processes (ie, the response inhibition process) but also leads to increased participant engagement [80,81]. However, the game must be validated against a basic task to ensure the efficacy of gamification and validity of measurement. In this paper, we present the design of a gamified SST (the stop-signal game [SSG]) and evaluate it relative to the basic task in 2 experiments that consider the effect of gamification on both performance and player experience.

Methods

Overall Procedure

This study sets out to evaluate a gamified version of the SST—termed SSG—along the 2 dimensions of performance and experience. We employed 2 studies to show that performance in a standard SST and in the new SSG was comparable within (study 1) and between (study 2) participants. Thus, comparing performance data in study 1 would give insight into the comparability of both tasks without adding unexplained variance in the form of interindividual variability. Study 2 aimed to replicate the results from study 1; a robust result should still hold even for between-group comparisons. Furthermore, we measured motivation and flow using the Intrinsic Motivation Inventory (IMI) [49] and the Flow State Scale (FSS) [82] in both studies to measure participant experience. Finally, in study 2, we employed an eye-tracking protocol to explain the influence of complex graphics on gaze behavior, and ultimately participant performance. We are of the opinion that the risk of sequence and carry-over effects in eye-tracking studies is especially high as participants pay increased attention to their eye movements. Thus, eye tracking was only employed in study 2. The whole eye-tracking procedure was comparable with earlier studies utilizing eye tracking in combination with the SST [83]. This paper aims to show that by leveraging gamification, cognitive tasks can be redesigned to produce more realistic and better data, without compromising internal validity [30,78].

Study 1: Within-Subject Design Participants

A total of 30 young, healthy adults were recruited for the study (16 female, 13 male, and 1 nonbinary). The mean age was 23.6 years (SD 4.51; range 17-35 years). The study was approved by the behavioral research ethics board of the University of Saskatchewan. All participants provided written informed consent.

Power Analysis

A power analysis was carried out using G.Power 3.1.3 [84]. For a medium-sized effect (η^2=0.15 or f=0.42), a medium-sized correlation between repeated measures, a power of 1–β=0.95, and an α value of .05, a minimum sample of 22 participants was needed to detect a significant difference between performance or subjective experience. Thus, failure to find a significant effect will support the null hypothesis.

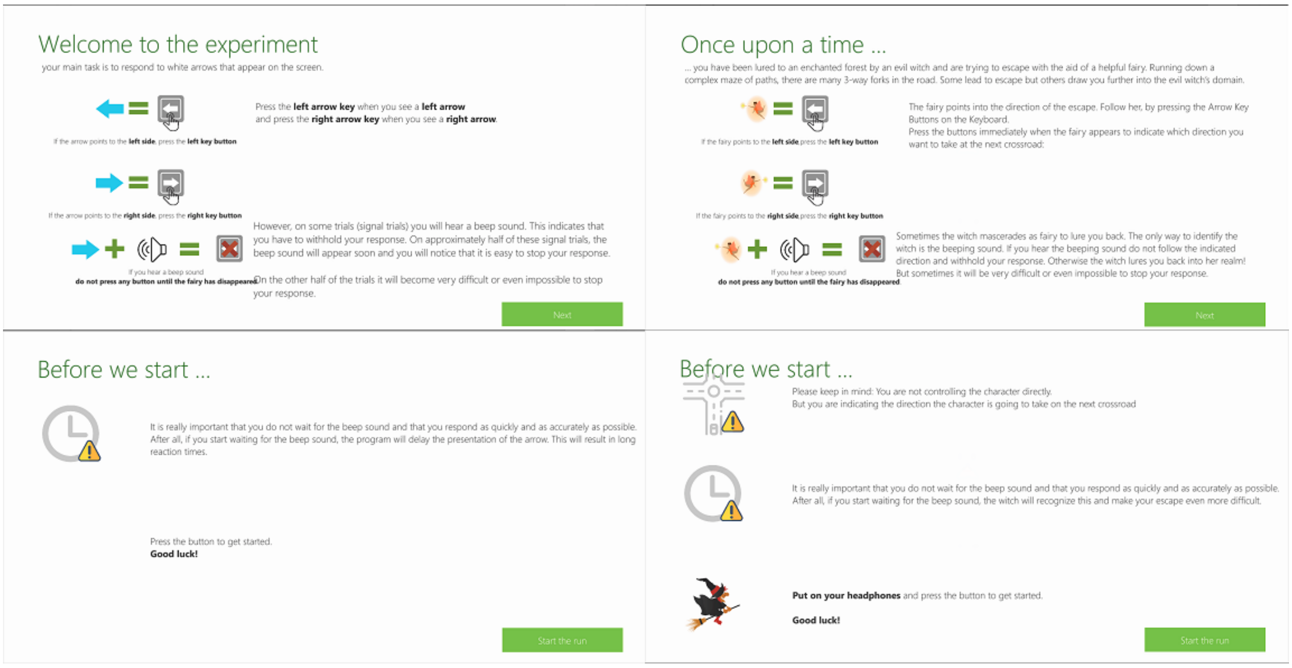
SSG Design

Both the SST and SSG were implemented using the Unity3D engine. The basic SST as well as the SSG consisted of 3 blocks, each containing 100 trials, 75% of which were go-trials and 25%, stop-trials. Between separate blocks, a pause of 15 seconds was granted. The go-stimulus was presented for a maximum of 1500 msec or until reaction. The stop signal was played over headphones following a variable delay (SSD), which was initially set to 250 msec. The SSD was continuously adjusted with the staircase procedure to obtain a probability of responding to 50%. After the reaction was successfully stopped (ie, button press was inhibited), the SSD was increased by 50 msec, whereas when the participants did not stop successfully, the SSD was decreased by 50 msec. The intertrial interval was set to a random value between 500 msec and 1500 msec. In the basic SST, participants had to respond to a left- or right-pointing arrow, which was presented in the upper third portion of the screen (Figure 1).

Figure 1. The stop-signal task (left) vs stop-signal game (right) trial appearance.

Our goals in designing the SSG were to make the game as identical to the task as possible, while also providing enjoyment. As many gamified cognitive tasks end up being experienced as disappointing in terms of enjoyment [24,28], we built our game around a popular game genre and ensured professional quality graphics. The SSG was built on the 3D infinite runner genre, in which the player sees a third-person view of their avatar running down a path (similar to the popular mobile game Temple Run by Imangi Studios). The game premise was integrated with the task instructions, as shown in Figure 2: “Once upon a time you have been lured to an enchanted forest by an evil witch are trying to escape with the aid of a helpful fairy.” In contrast to the arrows used in the SST, the SSG presented arrows in the form of a magical fairy who was pointing to the left or right and would guide them out of the forest (Figure 1). However, players were told that the evil witch sometimes masqueraded as the fairy, and the only way to know was through a beeping sound (ie, the auditory stop signal); in this case, they were to withhold their response or be lured deeper into the forest. After a choice was made, the avatar turned in the direction selected by the player, regardless of whether or not it was correct (technically, the camera rotated the world and the avatar continued straight). If they failed to respond or correctly withheld their response, the avatar continued on straight. Each choice occurred at a crossroad so that all options were possible, regardless of player response. The terrain was procedurally generated and shaded so that the forest was very dark (matching the dark background of the SST). As shown in Figure 1, we used “low poly” game art, which refers to meshes in 3D computer graphics that contain a small number of polygons, to give a professional appearance in real time apps (ie, games), while optimizing performance.

Figure 2. Instructions shown to participants for the stop-signal task (left) and stop-signal game (right).



The task was implemented using the Game Engine by Unity3D (version 2019.01) in a single version with a toggle button to switch between the SST and SSG (to keep the implementation of the underlying task identical). The differences between the SST and the SSG were the inclusion of a narrative theme and premise and the presence of the graphical elements, which included the background, the player avatar, and the stimulus (arrow or pointing fairy). The pointing fairy was designed to make the direction easily discriminable to avoid effects from overhead movement interfering with processing the intended movement direction, as replacing basic stimuli with more visually complex ones has been suggested to influence cognitive task performance [28]. In terms of gamification elements employed, we did not use points, scores, or a win or loss condition in the SSG but employed narrative elements including a backstory, a theme, and characters along with immersive elements of a 3D world and theme-appropriate graphics.

Stimuli and Apparatus

Participants were seated in front of a 27-inch color monitor with a viewing distance of approximately 80 cm. The study took place in an ordinarily lit room, 1 participant at a time. Participants were tasked to respond to the signals on screen by using 2 marked keys on the keyboard and withhold their response when an auditory stop signal (900 Hz) was presented over headphones. Participants were instructed to react as fast and accurately as possible. They were tasked to complete both SSTs: the basic SST and the SSG, as previously described. Both SST and SSG took approximately 15 min to complete.

Questionnaire Measures

A total of 2 established questionnaires were used to assess participant experience. The IMI measures motivation on 4 different subscales: interest-enjoyment, effort-importance, perceived competence, and tension-pressure. Each item was rated through an agreement with a statement on a 7-point scale (higher=greater agreement). The FSS assesses the subjective flow experience and factors influencing it using 9 subscales: challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on the task at hand, paradox of control, loss of self-consciousness, transformation of time, and autotelic experience. The FSS items were measured through an agreement with statements on a 5-point scale (higher=greater agreement).

Procedure

Participants were tasked to complete both the basic SST and the SSG, each followed immediately by the FSS and the IMI. The order of presentation of the SST and SSG was counterbalanced across participants and included as a factor in the analysis. After completion of both tasks and questionnaire sets, participants completed a demographic questionnaire.

Design

The study was based on a 2 (task: SST, SSG) x 2 (task-order: SST-SSG vs SSG-SST) mixed measures design with task as a within-subjects factor and task-order as a between-subjects factor.

Data Exclusion

In the data reduction phase, participants were excluded if they were uncooperative or produced faulty data. Initially, it was checked that all participants had normal or corrected-to-normal vision and hearing. For participant exclusion based on the SST and SSG performance, we followed the recommendations in the literature [85,86] that the SSRT can be reliably estimated for each participant in both sessions. Specifically, p(response|signal) had to be .4-.6, the horse-race model had to be satisfied, and the participant should not display strategic behavior (eg, waiting for the stop signal to appear). Furthermore, outliers based on the Tukey outlier criterion [87] within the data were identified and removed if necessary. After these procedures, 6 participants were excluded resulting in a sample of 24 with valid behavioral data. Furthermore, 1 additional participant had to be removed from the questionnaire analysis owing to a data collection error.

Dependent Measures

The main dependent variable for performance was the SSRT, that is, the estimate of time needed to respond to the stop signal and to cancel the movement, which measures the covert inhibition process. The estimation of the SSRT was based on the integration method with the replacement of omissions [85,88]. We also measured the SSD, the overall reaction time (RT) for both signal and no-signal trials, the probability of correct inhibition (p(response|signal)), and the omission and commission errors, as standard measures within the stop-signal paradigm [85]. The main dependent variables for experience were measured by using the IMI and FSS, as previously described.

Hypotheses

We hypothesized that there would be no difference in performance measures between the SSG and SST. Questionnaire data were analyzed to test our hypothesis that the SSG would elicit a more positive subjective experience as compared with the basic SST in terms of motivation and flow.

Study 2: Between-Subject Design

Sample

A total of 39 healthy subjects (20 female and 19 male) aged between 18 and 36 years (mean age 24.26, SD 4.99 years) were recruited for the study. All participants had normal or corrected-to-normal hearing and vision. The study was approved by the behavioral research ethics board of the University of Saskatchewan. All participants provided written informed consent.

Eye Tracking

We used a 60 Hz Tobii 4C eye tracker to measure the user's gaze focus. Areas of interest (AOIs) were mapped inside the app for subsequent analysis. The most important AOI was the instruction location (ie, stop-and-go signal location). For the SSG, additional AOIs were defined, including the path, the avatar, and the background.

Stimuli and Apparatus

These were identical as in study 1, apart from the eye-tracking device, which was mounted below the monitor.

Questionnaires

These were identical as in study 1.

Procedure

Upon entering the laboratory, participants were randomly divided into 2 groups: (1) basic SST and (2) gamified SSG. The eye tracker was calibrated for each participant; after calibration, participants started with the assigned task. After task completion, participants completed the questionnaires.

Design

The experiment was based on a two-group design. Each group was tasked to only complete either the basic SST or the gamified SSG.

Data Analysis

The experiment was based on a two-group (task: SSG vs SST) design. All other details are identical to study 1.

Data Exclusion

The procedure was identical to study 1. A total of 9 participants had to be excluded during the data reduction process resulting in a final sample of 30, evenly split between the 2 groups.

Results

Overview

A summary of the inference statistics in table form can be found in Multimedia Appendix 1. The tables (A1; stop-signal measures), A2 (IMI), and A3 (Flow) show means and SDs of measures in study 1, whereas tables A4 (stop-signal measures), A5 (IMI), and A6 (Flow) show the means and SDs for study 2.

Study 1: Within-Subject Design

Performance Results

Control Analysis

It is recommended to validate the obtained stop-signal data by showing a significant difference between the average signal RT and the average no-signal RT, with higher RTs for no-signal trials. To this end, a 2 (task-order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) x 2 (trial-type: signal vs no-signal) multivariate analysis of variance (MANOVA) was calculated. Only the main effect trial type ($F_{1,22}$=300.38; P<.001; η^2=0.92) was statistically significant, which shows that signal RT and no-signal RT were different in the expected direction. The main effect task ($F_{1,22}$=2.55; P=.13) and the main effect order ($F_{1,22}$=1.28; P=.27) were nonsignificant. Furthermore, all the two-way interactions and the three-way interactions did not yield a statistically significant result (all F<1).

SSRT

SSRT is an indirect estimate for the duration of the cognitive inhibition process, in which lower values represent higher inhibition speeds and efficiency. SSRTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task type ($F_{1,22}$=0.03; P=.86) and the main effect order ($F_{1,22}$=0.55; P=.47) as well as the interaction ($F_{1,22}$=0.02; P=.88) were not significant. To illustrate the comparable SSRT values for both task types, Figure 3 shows the SSRT distribution depending on the task type. Thus, the speed of the inhibition process was not altered by any experimental manipulation, providing support for the equivalence of the 2 task types concerning their measurement properties (Table 1).

Figure 3. Stop-signal reaction time (in milliseconds) distribution depending on task type for study 1.

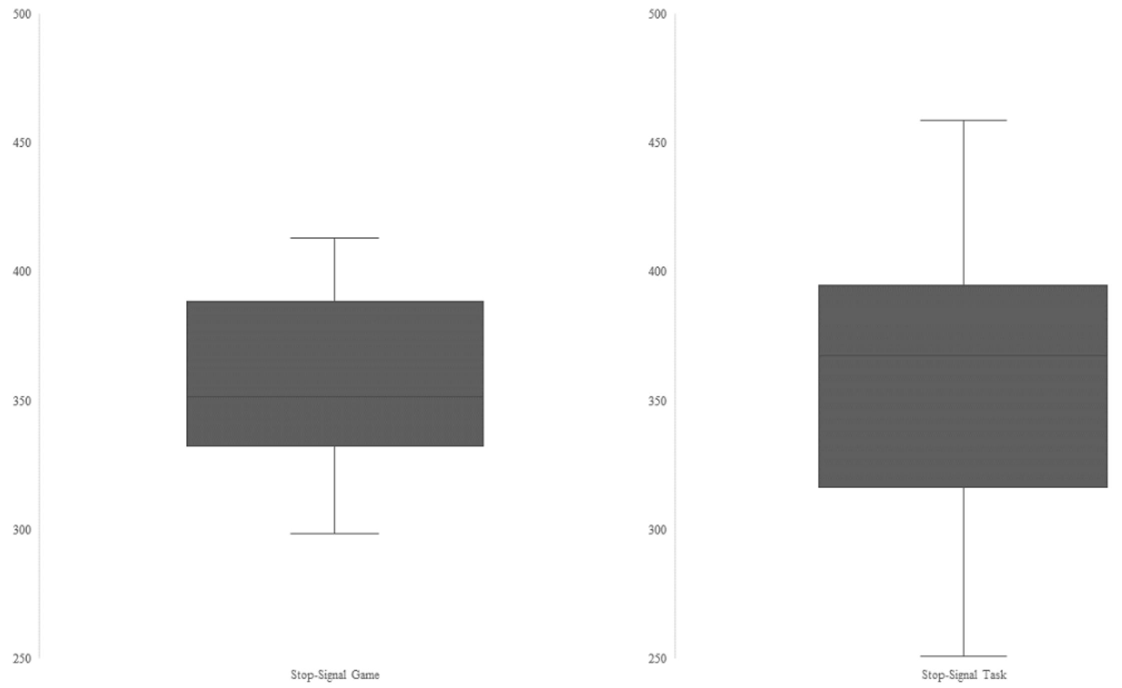


Table 1. Mean reaction time in milliseconds dependent on task type for study 1. The measurements are collapsed across the order of task performance.

Task	Signal reaction time	No-signal reaction time	Stop-signal delay	Stop-signal reaction time
Stop-signal task, mean (SD)	714.69 (164.83)	817.90 (181.60)	445.97 (185.87)	360.03 (53.86)
Stop-signal game, mean (SD)	750.67 (166.65)	851.55 (181.17)	476.02 (179.39)	358.33 (33.28)

SSD

A 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA was computed. The main effect task type ($F_{1,22}$=1.28; P=.27) and the main effect order ($F_{1,22}$=1.51; P=.23) as well as the interaction ($F_{1,22}$=0.02; P=.89) were not significant. Thus, there was no performance difference on the SSD depending on order or the type of SST (Table 1).

Signal RT

The incorrect signal RTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. Neither the main effect task ($F_{1,22}$=2.68; P=.12) nor the main effect order ($F_{1,22}$=1.59; P=.22) or the interaction ($F_{1,22}$=0.26; P=.31) were significant. This indicates that signal RT was not dependent on order or task type (Table 1).

No-Signal RT

Correct no-signal RTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task ($F_{1,22}$=2.18; P=.15), the main effect order ($F_{1,22}$=1.02; P=.33), and the interaction ($F_{1,22}$=0.004; P=.95) were all not significant. This result illustrates that overall correct RTs were not dependent on order or task type (Table 1).

Correct Inhibition

The probability of correctly inhibiting a response (p(response|signal)) was analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task type ($F_{1,22}$=0.01; P=.93) and the main effect order ($F_{1,22}$=1.10; P=.31) as well as the interaction ($F_{1,22}$=0.01; P=.93) were not significant. Thus, there was no performance difference in correct inhibition depending on order or the type of SST employed (Table 2).

Table 2. Mean error rates and accuracy in their relative proportion to the total trial count dependent on task type for study 1. The measurements are collapsed across the order of task performance.

Task	p(response|signal)	Omission error	Choice error
Stop-signal task, mean (SD)	0.47 (0.03)	0.04 (0.074)	0.0043 (0.007)
Stop-signal game, mean (SD)	0.47 (0.03)	0.016 (0.023)	0.0024 (0.0046)

Error Analysis

A total of 2 types of errors can be made during go-trials: omission errors (ie, missing a response) and commission errors (ie, choosing the wrong directional reaction). Both were analyzed using 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVAs. For omission errors, the main effects task ($F_{1,22}$=3.40; P=.08), the main effect order ($F_{1,22}$=0.10; P=.76) and the two-way interaction ($F_{1,22}$=0.61; P=.44) were not significant. Similarly, the main effect task ($F_{1,22}$=2.85; P=.11) and the main effect order ($F_{1,22}$=0.69; P=.42) and the interaction ($F_{1,22}$=0.22; P=.64) were not significant with regard to the commission errors. Taken together, order and task type did not influence error rates (Table 2).

Experience Results

IMI

As a first step, reliability scores for all 4 IMI subscales were calculated, using Cronbach alpha. The 4 subscales interest-enjoyment, perceived competence, effort-importance, and tension-pressure showed reliability scores of α_{ie}=.90, α_{pc}=.91, α_{ei}=.82, and α_{tp}=.70, which were deemed satisfactory. As a second step, all scores were analyzed using separate 2 (task: SSG vs SST) x 2 (order: SST-SSG vs SSG-SST) MANOVAs. For the subscale interest-enjoyment, a significant main effect of task was observed ($F_{1,21}$=16.35; P=.001; η^2=0.44), whereas the main effect order ($F_{1,21}$=0.03; P=.88) and the two-way interaction ($F_{1,21}$=0.03; P=.86) were not significant. In detail, participants rated interest-enjoyment on average 0.8 points higher (SD 0.914) for the SSG compared with the SST. The type of task did not affect ratings on all other subscales. For perceived competence, the main effect task ($F_{1,21}$=0.69; P=.41), the main effect order ($F_{1,21}$=0.56; P=.46), and the two-way interaction ($F_{1,21}$=0.81; P=.38) were not significant. For the subscale effort-importance, the main effect task ($F_{1,21}$=0.02; P=.90) and the main effect order ($F_{1,21}$=0.004; P=.95) as well as the interaction ($F_{1,21}$=0.02; P=.90) were not statistically significant. Finally, the subscale tension-pressure was not modulated by task ($F_{1,21}$=0.71; P=.71), order ($F_{1,21}$=0.85; P=.37), or the interaction between the 2 variables ($F_{1,21}$=1.10; P=.31). In summary, participants rated the game higher in interest-enjoyment compared with the basic version; the order in which tasks were completed did not affect the results. Owing to a lack of an overall scale score, the between-task difference values (Δ SSG-SST) were submitted to a multivariate analysis to determine whether or not overall IMI ratings differed among the tasks. The analysis revealed that when considering all subscales simultaneously, the SSG scored significantly higher compared with the SST ($F_{4,18}$=6.35; P=.002; η^2=0.59). Taken together, the analysis shows that the SSG scored significantly higher on the subscale interest-enjoyment (Cohen d=0.601) and was overall rated higher on the IMI (Cohen d=1.109). For scale means, refer to Table 3.

Table 3. Mean scale values for each Intrinsic Motivation Inventory subscale depending on task variant and the study (study 1).

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Interest-enjoyment	3.25 (1.45)	4.05 (1.20)
Perceived competence	3.83 (1.39)	3.99 (1.41)
Effort-importance	4.68 (1.39)	4.71 (1.32)
Tension-pressure	3.29 (1.17)	3.49 (1.11)

Flow

As a first step, reliability scores for all 9 flow subscales and the complete scale were calculated using the Cronbach alpha. The 9 subscales, challenge-skill balance (α_{csb}=.85), action-awareness merging (α_{awm}=.78), clear goals (α_{cg}=.86), unambiguous feedback (α_{uf}=.81), concentration on the task at hand (α_{c}=.66), paradox of control (α_{pc}=.90), loss of self-consciousness (α_{lsc}=.82), transformation of time (α_{tt}=.82), autotelic experience (α_{ae}=.87), and the overall scale ($\alpha_{overall}$=.91), showed satisfactory reliability scores. As a second step, all scores were analyzed using separate 2 (task: SSG vs SST) x 2 (order: SST-SSG vs SSG-SST) MANOVAs. For the subscales challenge-skill balance, action-awareness merging, clear goals, paradox of control, loss of self-consciousness, and transformation of time, no significant effects emerged. In detail, with regard to the action-awareness merging subscale, the main effect task ($F_{1,21}$=3.38; P=.08), the main effect order ($F_{1,21}$=0.85; P=.36), and the interaction ($F_{1,21}$=0.30; P=.59) were not significant. The main effects task ($F_{1,21}$=0.48; P=.50), order ($F_{1,21}$=0.04; P=.85), and their interaction ($F_{1,21}$=0.001; P=.98) were not significant with regard to the action-awareness merging subscale. The analysis of the subscale clear goals neither revealed a significant main effect task ($F_{1,21}$=2.24; P=.15) nor a main effect order ($F_{1,21}$=0.06; P=.81) and no interaction ($F_{1,21}$=0.16; P=.69). Neither the task type ($F_{1,21}$=3.83; P=.06) nor the order ($F_{1,21}$=0.32; P=.58,) or the interaction between task x order ($F_{1,21}$=2.06; P=.17) were significant for the subscale paradox of control. Loss of self-consciousness was not modulated by the task ($F_{1,21}$=1.72; P=.20) or by the order ($F_{1,21}$=0.10; P=.76), and there was no interaction between the 2 variables ($F_{1,21}$=2.73; P=.11). The ratings for transformation of time were neither influenced by the task ($F_{1,21}$=0.14; P=.71) nor by the order ($F_{1,21}$=0.002; P=.96) or their interaction ($F_{1,21}$=0.02; P=.91). All effects with regard to the subscale unambiguous feedback were significant. In detail, the main effect task ($F_{1,21}$=5.76; P=.04; η^2=0.22) and the interaction of task x order ($F_{1,21}$=5.76; P=.04; η^2=0.22) displayed equally large effects whereas the effect for the main effect order ($F_{1,21}$=3.77; P=.06; η^2=0.15) was slightly smaller. Taken together, these results show that unambiguous feedback was rated higher in the game version than the basic task and this effect was enhanced when participants first worked on the basic version and then played the game. The interaction of task x order for the concentration on the task at hand score was significant ($F_{1,21}$=6.81; P=.02; η²=0.25) whereas the main effect task ($F_{1,21}$=1.06; P=.31) and order ($F_{1,21}$=0.11; P=.74) were not, showing that concentration decreased in the second session regardless of which task version was done first or second. For an autotelic experience, the main effect task ($F_{1,21}$=6.79; P=.02; η²=0.24) was significant whereas the main effect order ($F_{1,21}$=0.40; P=.53) and the interaction ($F_{1,21}$=0.0002; P=.99) were not, showing that participants were more internally driven playing the game version over the basic version. Most importantly, the overall Flow scale score was significantly higher for the SSG compared with the SST as indicated by the main effect task ($F_{1,21}$=5.92; P=.02; η²=0.22) and there was no main effect order ($F_{1,21}$=0.54; P=.47) or an interaction between task x order ($F_{1,21}$=0.14; P=.71). To summarize, concentration on the task at hand decreased in the second session, which can be attributed to fatigue. In addition, the experience of unambiguous feedback increased when participants did the basic task first and then played the game. This result illustrates the participants' feelings that the performance feedback was better and more responsive in the game version compared with the basic task. Furthermore, the results show that the overall experience of flow and the autotelic experience in particular were rated higher in the gamified version of the task. For mean values of the scale, refer to Table 4.

Table 4. Mean scale values for each Flow subscale depending on the task variant for study 1.

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Challenge-skill balance	3.10 (.90)	3.28 (.81)
Action-awareness merging	3.38 (.75)	3.26 (.84)
Clear goals	3.89 (.71)	4.03 (.62)
Unambiguous feedback	3.37 (.73)	3.52 (.83)
Concentration on the task	3.04 (.80)	3.26 (.75)
Paradox of control	3.12 (.96)	3.32 (.88)
Loss of self-consciousness	3.16 (1.03)	3.35 (1.00)
Transformation of time	3.05 (1.01)	3.00 (.86)
Autotelic experience	2.43 (.84)	2.78 (.79)
Overall	3.17 (.54)	3.31 (.49)

Bayesian Analysis

We employed the Bayesian analysis to put our results to an additional test and support any eventual interpretation of our data. Task version difference scores were calculated for each dependent variable (ie, all performance measures and scale values) to reflect the difference between SST and SSG. The difference scores for all performance variables (eg, SSRT, interest-enjoyment, challenge-skill balance) were submitted to Bayesian paired sample t tests using JASP. For performance measures, two-tailed tests were used, and for questionnaires, one-tailed tests were used. We used a Cauchy prior distribution with r=0.707. This prior was chosen because it reflects the range of most psychological effects [89] but given our hypothesis of a nonsignificant difference between the 2 task variations, it is a somewhat conservative prior. For the behavioral performance measures, an unspecific alternate hypothesis was specified (H1: SST≠SSG), whereas for the questionnaire data, a hypothesis-conform alternate hypothesis was chosen (H1: SSG>SST). The Bayesian analysis showed that there were no performance differences between SST and SSG. In detail, results showed weak-to-moderate support for the null hypothesis, and H0 was up to 4.63 times as likely as the alternative hypothesis depending on the behavioral performance measure in question [90-92]. With regard to the IMI, the analysis revealed decisive evidence for the subscale interest-enjoyment (BF_{10}=168.11), whereas for all other IMI subscales, the null hypothesis was more likely (BF_{01} ranging from 2.4 to 4.19). Flow analysis showed mostly indecisive BFs, but there was moderate support for H0 with regard to the 2 subscales, action-awareness merging (BF_{01}=7.21) and transformation of time (BF_{01}=5.91). In contrast, the analysis revealed moderate support for H1 with regard to the autotelic experience (BF_{10}=7.61) and the overall flow experience (BF_{10}=4.91; Table 5).

Table 5. The Bayes factor table for study 1 shown by BF01 and BF10.

Dependent measures	Study 1	
	BF_{01}[a]	BF_{10}
Behavioral Performance Measures[b]		
No-signal reaction time	1.72	0.58
Signal reaction time	1.48	0.68
Stop-signal reaction time	4.60	0.22
Stop-signal delay	2.52	0.40
Omission errors	1.16	0.86
Choice errors	1.37	0.73
p(response|signal)	4.63	0.22
Intrinsic Motivation Inventory		
Interest-enjoyment	0.006	168.11
Perceived competence	2.40	0.42
Effort-importance	4.19	0.24
Tension-pressure	2.44	0.41
Flow State Scale		
Challenge-skill balance	0.60	1.68
Action-awareness merging	7.21	0.14
Clear goals	0.82	1.22
Unambiguous feedback	0.51	1.96
Concentration	1.34	0.75
Paradox of control	0.66	1.52
Loss of self-consciousness	1.65	0.61
Transformation of time	5.91	0.17
Autotelic experience	0.13	7.61
Overall	0.20	4.92

[a]BF: Bayes factor.

[b]For the behavioral performance measures H0: SSG=SST and H1: SSG≠SST. For questionnaire measures H1: SSG>SST.

Study 2: Between-Subject Design

Performance Measures

Control Analysis

To establish that signal RT and no-signal RT significantly differ from each other, a 2 (task: SSG vs SST) x 2 (trial-type: signal vs no-signal) MANOVA was calculated. Only the main effect trial type ($F_{1,28}$=156.14; P<.001; η^2=0.85) was statistically significant, which showed that signal RT and no-signal RT trials were different in the expected direction. The main effect of task ($F_{1,28}$=0.14; P=.71) as well as the two-way interaction ($F_{1,28}$=1.37; P=.25) were not significant.

SSD

There was no significant difference between the game and the basic version with regard to SSD ($F_{1,28}$=0.00006; P=.99).

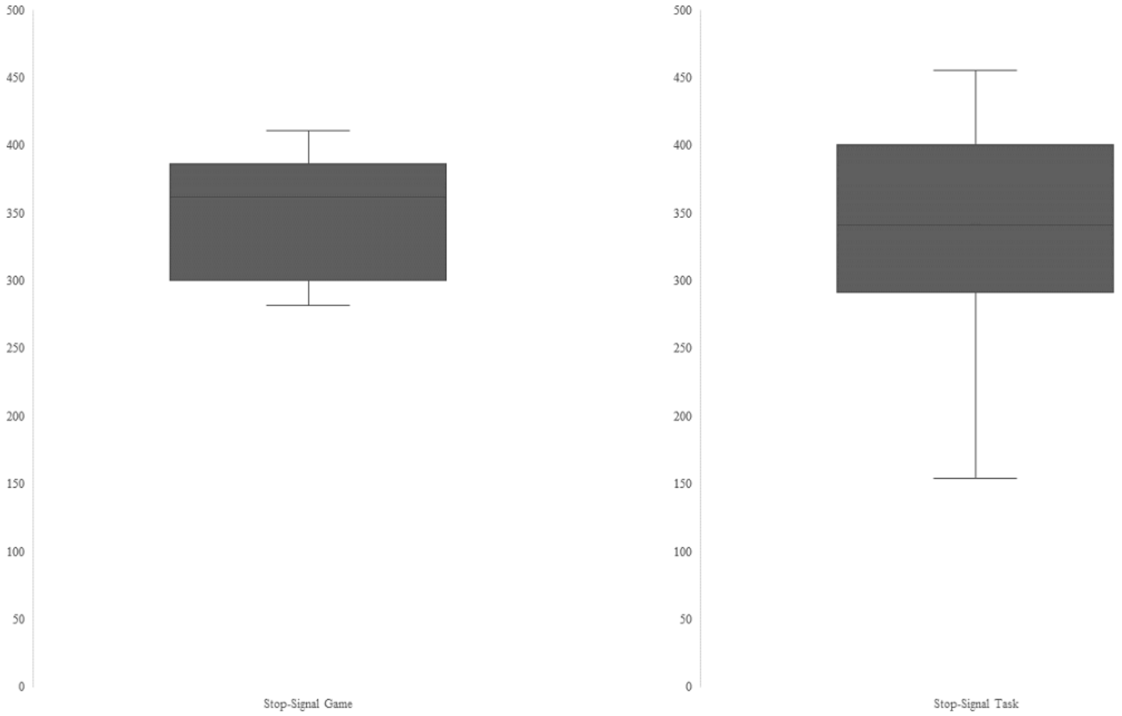
SSRT

A one-way between-subjects analysis of variance (ANOVA) was conducted to compare the effect of task type (SSG vs SST) on SSRT. The main effect task ($F_{1,28}$=0.03; P=.87) was not significant. Thus, the speed of the inhibition process did not depend on the task (Table 6; Figure 4).

Table 6. Mean reaction time in milliseconds dependent on the task type for study 2.

Task	Signal reaction time	No-signal reaction time	Stop-signal delay	Stop-signal reaction time
Stop-signal task, mean (SD)	648.12 (137.99)	759.90 (171.10)	406.03 (207.00)	344.33 (75.73)
Stop-signal game, mean (SD)	681.49 (183.49)	774.10 (204.26)	405.43 (208.52)	348.08 (46.11)

Figure 4. Stop-signal reaction time (in milliseconds) distribution depending on task type for study 2.



Signal RT

A one-way between-subjects ANOVA showed no difference between the game and basic version of the SST ($F_{1,28}$=0.32; *P*=.58; Table 6).

No-Signal RT

Correct no-signal RTs did not differ between the task types ($F_{1,28}$=0.04; *P*=.84; Table 6).

Correct Inhibition

The probability of correctly inhibiting a response (p(response|signal)) did not differ between the basic and game version ($F_{1,28}$=0.79; *P*=.38; Table 7).

Table 7. Mean error rates and accuracy in their relative proportion to the total trial count dependent on the study and task type for study 2.

Task	p(response|signal)	Omission error	Choice error
Stop-signal task, mean (SD)	.49 (.04)	0.019 (0.034)	0.0021 (0.0038)
Stop-signal game, mean (SD)	.47 (.04)	0.020 (0.035)	0.0033 (0.0043)

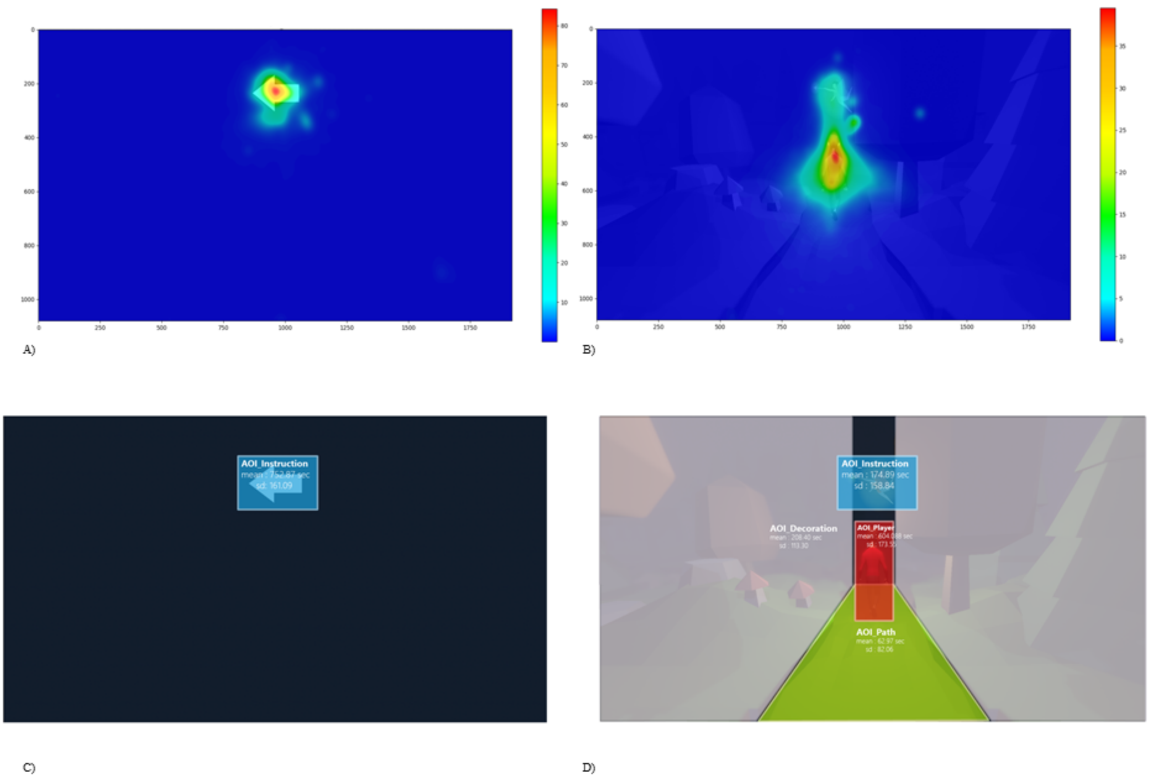
Error Analysis

Neither the omission errors ($F_{1,28}$=0.005; *P*=.94) nor the commission errors ($F_{1,28}$=0.66; *P*=.42) differed between the 2 task versions. Taken together, order and task type did not influence error rates (Table 7).

Eye Tracking

We recorded the estimated gaze fixation per user, which is the average fixation of the 2 eyes of the user. This screen coordinate was mapped on the previously introduced AOIs. Per user, we calculated the average focused time per AOI for both conditions over the complete experiment. In the gamified condition, users mostly focused on the avatar (mean 604.09 seconds, SD 173.55), less on the environment (mean 208.41 seconds, SD 113.31), and least on the instruction location (mean 174.89 seconds, SD 158.84), whereas they mostly looked at the instruction location in the basic version of the task (mean 752.88, SD 161.09). For an illustration of the results, see Figure 5.

Figure 5. Visualization of the eye-tracking results. Parts (A) and (B) display the gaze focus as a heat map. Parts (C) and (D) display the mean time that the participants spent focused on parts of the display.



Experience Measures

IMI

The reliability for all subscales was calculated. The 4 subscales interest-enjoyment, perceived competence, effort-importance, and tension-pressure showed reliability scores of α_{ie}=.87, α_{pc}=.89, α_{ei}=.77, and α_{tp}=.81, respectively. All subscale scores were submitted to a one-way between-subject ANOVA comparing the basic and the gamified task version. There were no significant differences for interest-enjoyment ($F_{1,28}$=0.001; P=.98), perceived competence ($F_{1,28}$=0.28; P=.60), effort-importance ($F_{1,28}$=1.28; P=.27) and tension-pressure ($F_{1,28}$=0.57; P=.46; Table 8).

Table 8. Mean scale values for each Intrinsic Motivation Inventory subscale depending on task variant and the study.

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Interest-enjoyment	3.09 (1.21)	3.11 (1.29)
Perceived competence	3.6 (1.29)	3.84 (1.19)
Effort-importance	4.3 (1.39)	4.78 (.89)
Tension-pressure	3.43 (1.44)	3.80 (1.22)

Flow

The reliabilities for all subscales and the overall reliability for the Flow scale was calculated. In detail, the 9 subscales, challenge-skill balance (α_{csb}=.66), action-awareness merging (α_{awm}=.66), clear goals (α_{cg}=.71), unambiguous feedback (α_{uf}=.74), concentration on the task at hand (α_{c}=.82), paradox of control (α_{pc}=.86), loss of self-consciousness (α_{lsc}=.71), transformation of time (α_{tt}=.79), autotelic experience (α_{ae}=.88), and the overall scale ($\alpha_{overall}$=.86), showed satisfactory reliability scores. Overall, there were no statistical differences between the game and the basic version. In detail, the subscales for challenge-skill balance ($F_{1,28}$=0.05; P=.83), action-awareness merging ($F_{1,28}$=0.07; P=.80), clear goals ($F_{1,28}$=3.20; P=.08), unambiguous feedback ($F_{1,28}$=0.72; P=.40), the concentration on the task at hand ($F_{1,28}$=3.27; P=.08), paradox of control ($F_{1,28}$=1.36; P=.25), loss of self-consciousness ($F_{1,28}$=1.56; P=.22), transformation of time ($F_{1,28}$=0.65; P=.43), and autotelic experience ($F_{1,28}$=0.01; P=.91) as well as all subscales combined ($F_{1,28}$=0.83; P=.37) did not differ between the basic and game version (Table 9).

Table 9. Mean scale values for each Flow subscale depending on task variant and the study (study 2).

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Challenge-skill balance	3.08 (7.11)	3.13 (.52)
Action-awareness merging	3.20 (.71)	3.27 (.71)
Clear goals	3.93 (.50)	3.60 (.52)
Unambiguous feedback	3.47 (.77)	3.27 (.79)
Concentration on the task	3.17 (.78)	2.67 (.73)
Paradox of control	2.95 (.83)	3.28 (.73)
Loss of self-consciousness	3.53 (.92)	3.13 (.83)
Transformation of time	3.42 (.83)	3.18 (.75)
Autotelic experience	2.28 (.96)	2.32 (.59)
Overall	3.23 (.44)	3.09 (.34)

Bayesian Analysis

Similar to study 1, we tested the 2 stopping task types against each other using a Bayesian independent sample *t* test with the same parameters as in study 1. For the behavioral performance measures, an unspecific alternate hypothesis was specified (H1: SST≠SSG), whereas for the questionnaire data, a hypothesis-conform alternate hypothesis was chosen (H1: SSG>SST). We obtained moderate evidence for the null hypothesis with regard to the behavioral performance measures, confirming that there is no performance difference between the two. The analysis of IMI scores revealed no conclusive BFs and only a tendency toward the null hypothesis. Results of the FSS analysis showed moderate evidence for the null hypothesis in several subscales (ie, clear goals, unambiguous feedback, concentration, loss of self-consciousness, and transformation of time) as well as the overall scale score (Table 10).

Table 10. The Bayes factor (BF) table showing BF01 and BF10.

Dependent measures	Study 2	
	BF_{01}[a]	BF_{10}
Behavioral Performance Measures [b]		
No-signal reaction time	2.86	.35
Signal reaction time	2.57	.39
Stop-signal reaction time	2.87	.35
Stop-signal delay	2.90	.34
Omission errors	2.90	.35
Choice errors	2.26	.44
p(response|signal)	2.15	.47
Intrinsic Motivation Inventory		
Interest-enjoyment	2.85	.35
Perceived competence	1.94	.52
Effort-importance	1.08	.93
Tension-pressure	1.58	.63
Flow State Scale		
Challenge-skill balance	2.48	.40
Action-awareness merging	2.41	.42
Clear goals	6.83	.15
Unambiguous feedback	4.73	.21
Concentration	6.78	.15
Paradox of control	1.04	.97
Loss of self-consciousness	5.63	.18
Transformation of time	4.63	.22
Autotelic experience	2.68	.37
Overall	4.87	.21

[a]BF: Bayes factor.

[b]For the behavioral performance measures H0: SSG=SST and H1: SSG≠SST. For questionnaire measures H1: SSG>SST.

Discussion

Principal Findings

Overall results show that our newly developed SSG can be used to measure response inhibition as well as SST while being more enjoyable. Specifically, in 2 studies employing a within-subject (study 1) and between-subject (study 2) design, we showed that there were no significant differences between the 2 tasks across all behavioral performance measures. Furthermore, we obtained strong evidence that the SSG was more enjoyable and led to higher experiences of flow but only when participants were able to compare the 2 tasks with each other.

In detail, the results of study 1 showed that performance did not differ between the SSG and the basic SST and that the order of tasks did not influence performance. Concerning the experience of flow and intrinsic motivation, the SSG was superior to the standard SST paradigm, with the largest effect being shown by the interest-enjoyment subscale in the IMI, in which 44% of the variance was explained by the game versus task manipulation. Importantly, effect sizes suggest the existence of a large difference between SST and SSG with regard to interest-enjoyment (η^2=0.44; Cohen *d*=0.601) and overall intrinsic motivation (η^2=0.59; Cohen *d*=1.109). Furthermore, the SSG scored higher on the flow subscales for an autotelic experience and unambiguous feedback, and the overall flow score was significantly higher for the SSG, with the game elements explaining 22% to 24% of the variance in experienced flow. These frequentist results were confirmed by the Bayesian analysis. First, there was evidence against performance differences between the 2 tasks. Second, we obtained decisive evidence for a higher interest-enjoyment rating in the SSG compared with the SST, whereas all other IMI subscales were not affected by the type of stopping task. Third, there was evidence for a higher level of autotelic and overall flow experience in the SSG compared with the SST. Overall, our findings suggest that the SSG can be used as a reliable measurement of the response inhibition process, while being experienced as more enjoyable for participants.

In a second study, we aimed to extend and replicate our findings. As there is evidence that stopping is influenced by perceptual distractors [83], we implemented an eye-tracking procedure to assess the gaze differences between the SST and the SSG. The eye-tracking implementation mirrors the exploratory analysis by Verbruggen et al [83]. Their exploratory analysis showed that the frequency of eye movements was increased in the condition where the stop signal was presented peripherally. If we had found a significant performance difference between the 2 task versions, we could have used the eye-tracking data to explain this result. On the contrary, our results show that despite a more visually complex environment, which modified gaze and eye movements, the SSG leads to a comparable performance with the SST. We opted for a between-subjects design, which has the additional benefit of eliminating any sequence effects on the eye-tracking data; especially, when people are aware that their behavior is tracked across different task versions, they might behave differently. With that being said, we expected the differences in subjective experience (ie, differences in questionnaire scores) to be smaller owing to the lack of a direct comparison in a between-subjects design.

The results of study 2 partially replicated the results of study 1. We did not find performance differences between SST and SSG in any performance measure. Contrary to study 1, an analysis of questionnaire scores showed that there was no difference between SST and SSG in either the IMI or Flow scale. A Bayes analysis confirmed these findings and revealed small-to-moderate evidence for the null hypothesis (H0: SST=SST) with regard to performance measures as well as the IMI and Flow scales. The lack of differences in questionnaire scores was somewhat expected. This result is likely due to the fact that participants in study 2 had no chance to compare the 2 task versions coupled with a regression toward the mean and a tendency of participants to avoid the more extreme scale ratings [93,94]. In addition, we think that the lack of a significant difference between tasks in study 2 is positive. It reflects that only when an implicit comparison between SST and SSG can be made are the 2 versions perceived differently, but overall, the influence of the gamification on motivation is not exceedingly large, as task performance was still comparable. Game elements that overly influence task performance can in turn make it difficult to gather an individual's exact baseline performance. The average fixation time on the previously presented AOIs showed that the gaze focus differed between the 2 tasks. However, this crucially did not seem to affect performance. Interestingly, this hints at the possibility that foveal focus and attention is not required to effectively process a simple stop signal.

Comparison With Previous Work

To the authors' knowledge, there only has been one other study that tried to gamify the SST [30]. The aforementioned study compared 3 SST variants—standard, theme, and scoring—in participants who were recruited and tested on the web. They found no effect of task variant on attrition, and although the variant with the scoring system had higher ratings, the theme variant scored lower compared with the standard SST paradigm. Importantly however there are several differences between the study by Lumsden et al [30] and this paper. We employed the SST and SSG in a controlled lab environment and not on the web, which has the clear advantage of control over the environment, the experimental set-up, and participant compliance. Furthermore, Lumsden et al [30] focused on the rounds played by participants after 4 required initial sessions but found no effect of task-variant playtime. Although the amount played might be a good measure of motivation, the reward for playing was low—monetarily and intrinsically.

In detail, the gamification used in the study by Lumsden et al [30] consisted of a scoring system without any graphical changes to the task or a thematic variation of the SST; in this case, the player had to sort fruit into different buckets. The theme version of the SST did not implement a scoring system, which is similar to the SSG in this study. Although this was not tested and is pure speculation, the authors of this study think it is reasonable to assume that the haunted forest cover story provides a higher sense of urgency and might be more engaging than sorting fruit by color. Furthermore, Lumsden et al [30] did award participants with only 50 cents for every session after the fourth session, which may not have been enough to keep players motivated. With that being said, the authors mirrored our results by showing that there were no performance differences between the task versions. We decided against the more volatile measure of play sessions and aimed to directly capture performance and motivation. Nevertheless, we think that the study provided important initial evidence and that it might be interesting in the future to validate our SSG on the web.

Limitations

This study has 3 important limitations. First, overall reaction times and inhibition speeds were elevated in both the SST and SSG compared with the ordinarily observed values [65,66,83,85,86,95]. Furthermore, there is evidence that SSRT is unaffected by the demands of the go task [95], but this is still up for debate [83,86]. However, RTs as reported in this study are not completely unusual, and as both tasks produced comparable performance measures, this elevation might be traced back to the samples. Second, we only found a reliable difference in flow and motivation between SSG and SST in study 1 (ie, within-subject design). As task-order did not affect the evaluation of SST or SSG, we speculate that the increased motivation and flow experience in the within-subject study was because both tasks could be compared side-by-side. This illustrates that those kinds of subjective questionnaire measures are somewhat context-dependent. Third, one could also take the neuroscientific approach to compare SST and SSG. In detail, if our claim is that the SSG is a methodologically valid and more enjoyable substitute of the SST that can measure response inhibition accurately, then similar neural correlates should be obtainable. Specifically, similar to the SST, we would expect the right prefrontal cortex to play a crucial role in stopping performance during the SSG and, in contrast to the SST, other areas more responsible for visual information processing should show increased activity during the SSG [65,66,96-100]. In addition, recent evidence suggests that performance differences in video games translate to differential brain activity [101].

Thus, future neuropsychological studies might have to take the individual baseline performance and brain activity into account.

Outlook

There are several directions in which future research could be taken. As already mentioned in the introductory section of this paper, the ability to inhibit an already initiated action is linked to mental health conditions such as ADHD, OCD, schizophrenia, and posttraumatic stress disorder [59,69,102,103]. As video games are accessible, motivating, and can be custom built to capture behavior, such as the SSG in this paper, it has been proposed that digital games or game-like tasks can be useful for the assessment and treatment of mental health issues [104,105]. Although the use of cognitive psychological testing in a clinical setting is well established and those experimental approaches produce reliable between-group (ie, clinical vs nonclinical sample) differences, they are not necessarily reliable on an individual level over time [106]. Nevertheless, we propose that the SSG should be experimented with its use in a more applied setting. This can be especially important in cases where obtaining a valid response inhibition measurement is difficult. For example, in some clinical subsamples, the ability to focus on the task at hand is limited, and the SSG is more easily accessible and motivating for participants although validly measuring the stopping ability. To this end, a first step could be to validate the present results on a larger scale in a remote web-based assessment. A web-based assessment of behavioral as well as psychophysiological measures via game-like tasks has been done before and shown to be promising for the future [107,108].

Conclusions

Taken together, our results suggest that the newly developed SSG is an effective tool to measure the response inhibition process. The SSG compared with the regular SST has 2 clear advantages. First, the SSG leads to higher enjoyment and flow, and second, it assesses an individual's stopping capabilities in a more realistic, ecologically valid setting.

Acknowledgments

The authors thank the Natural Sciences and Engineering Research Council of Canada and Saskatchewan-Waterloo Games User Research program for funding. They also thank Jeremy Storring, Parker Neufeld, Chase Crawford, and Katelyn Wiley, along with the rest of the members at the Interaction Lab at the University of Saskatchewan, and the participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the performance analysis results.
[DOCX File , 26 KB - games_v8i3e17810_app1.docx]

References

1. Deterding S, Dixon D, Khaled R, Nacke L. Privacy by Design: Informed Consent and Internet of Things for Smart Healthfrom Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; October 6-8, 2011; Tampere, Finland. [doi: 10.1145/2181037.2181040]
2. Robson K, Plangger K, Kietzmann JH, McCarthy I, Pitt L. Is it all a game? Understanding the principles of gamification. Bus Horiz 2015 Jul;58(4):411-420. [doi: 10.1016/j.bushor.2015.03.006]
3. Lee J, Hammer J. Gamification in education: what, how, why bother? Acad Exch Q 2011;15(2):146.
4. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User-Adap Inter 2014 Jul 14;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
5. Grund CK. How games and game elements facilitate learning and motivation: a literature review. Lect Notes Informatics 2015:- [FREE Full text]
6. Birk MV, Mandryk RL, Atkins C. The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY'16; October 21-23, 2016; Austin, Texas, USA. [doi: 10.1145/2967934.2968091]
7. David R, Michael L, Chen S. -. In: Serious Games: Games That Educate, Train, and Inform. New York, USA: Muska & Lipman/Premier-Trade; 2006.
8. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
9. Vaibhav A, Gupta P. Gamification of MOOCs for Increasing User Engagement. In: International Conference on MOOC, Innovation and Technology in Education. 2014 Presented at: MITE'14; December 19-20, 2014; Patiala, India. [doi: 10.1109/mite.2014.7020290]

10. Khalili-Mahani N, Assadi A, Li K, Mirgholami M, Rivard M, Benali H, et al. Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. JMIR Ment Health 2020 Mar 26;7(3):e12388 [FREE Full text] [doi: 10.2196/12388] [Medline: 32213474]
11. Zhou H, Fishbach A. The pitfall of experimenting on the web: how unattended selective attrition leads to surprising (yet false) research conclusions. J Pers Soc Psychol 2016 Oct;111(4):493-504. [doi: 10.1037/pspa0000056] [Medline: 27295328]
12. Bados A, Balaguer G, Saldaña C. The efficacy of cognitive-behavioral therapy and the problem of drop-out. J Clin Psychol 2007 Jun;63(6):585-592. [doi: 10.1002/jclp.20368] [Medline: 17457848]
13. Flick SN. Managing attrition in clinical research. Clin Psychol Rev 1988 Jan;8(5):499-515. [doi: 10.1016/0272-7358(88)90076-1]
14. Geraghty AW, Wood AM, Hyland ME. Attrition from self-directed interventions: investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. Soc Sci Med 2010 Jul;71(1):30-37. [doi: 10.1016/j.socscimed.2010.03.007] [Medline: 20400220]
15. Kirkwood MW, Kirk JW, Blaha RZ, Wilson P. Noncredible effort during pediatric neuropsychological exam: a case series and literature review. Child Neuropsychol 2010;16(6):604-618. [doi: 10.1080/09297049.2010.495059] [Medline: 20628928]
16. DeRight J, Jorgensen RS. I just want my research credit: frequency of suboptimal effort in a non-clinical healthy undergraduate sample. Clin Neuropsychol 2015;29(1):101-117. [doi: 10.1080/13854046.2014.989267] [Medline: 25494327]
17. Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, Conference Participants. American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. Clin Neuropsychol 2009 Sep;23(7):1093-1129. [doi: 10.1080/13854040903155063] [Medline: 19735055]
18. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. Behav Res Methods 2013 Jun;45(2):301-318. [doi: 10.3758/s13428-012-0264-3] [Medline: 23055169]
19. Bogacz R, Hu PT, Holmes PJ, Cohen JD. Do humans produce the speed-accuracy trade-off that maximizes reward rate? Q J Exp Psychol (Hove) 2010 May;63(5):863-891 [FREE Full text] [doi: 10.1080/17470210903091643] [Medline: 19746300]
20. Hinvest NS, Anderson IM. The effects of real versus hypothetical reward on delay and probability discounting. Q J Exp Psychol (Hove) 2010 Jun;63(6):1072-1084. [doi: 10.1080/17470210903276350] [Medline: 19890767]
21. Hickey C, Kaiser D, Peelen MV. Reward guides attention to object categories in real-world scenes. J Exp Psychol Gen 2015 Apr;144(2):264-273. [doi: 10.1037/a0038627] [Medline: 25559653]
22. le Pelley ME, Pearson D, Griffiths O, Beesley T. When goals conflict with values: counterproductive attentional and oculomotor capture by reward-related stimuli. J Exp Psychol Gen 2015 Feb;144(1):158-171. [doi: 10.1037/xge0000037] [Medline: 25420117]
23. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Apr 20;19(4):e128 [FREE Full text] [doi: 10.2196/jmir.6906] [Medline: 28428161]
24. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
25. Tong T, Chignell M, Lam P, Tierney MC, Lee J. Designing serious games for cognitive assessment of the elderly. Proc Int Symp Hum Factors Ergon Healh Care 2014 Jul 22;3(1):28-35. [doi: 10.1177/2327857914031004]
26. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: a scoping review. J Biomed Inform 2016 Dec;64:296-319 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.019] [Medline: 27815228]
27. Aeberhard A, Gschwind L, Kossowsky J, Luksys G, Papassotiropoulos A, de Quervain D, et al. Introducing COSMOS: a web platform for multimodal game-based psychological assessment geared towards open science practice. J Technol Behav Sci 2018 Sep 28;4(3):234-244. [doi: 10.1007/s41347-018-0071-5]
28. Wiley K, Vedress S, Mandryk L. How Points and Theme Affect Performance and Experience in a Gamified Cognitive Task. In: Conference on Human Factors in Computing Systems. 2020 Presented at: CHI'20; April 25-30, 2020; Oahu, Hawai i, USA. [doi: 10.1145/3313831.3376697]
29. Price RB, Kuckertz JM, Siegle GJ, Ladouceur CD, Silk JS, Ryan ND, et al. Empirical recommendations for improving the stability of the dot-probe task in clinical research. Psychol Assess 2015 Jun;27(2):365-376 [FREE Full text] [doi: 10.1037/pas0000036] [Medline: 25419646]
30. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M. Attrition from web-based cognitive testing: a repeated measures comparison of gamification techniques. J Med Internet Res 2017 Nov 22;19(11):e395 [FREE Full text] [doi: 10.2196/jmir.8473] [Medline: 29167090]
31. Lumsden J, Skinner A, Woods AT, Lawrence NS, Munafò M. The effects of gamelike features and test location on cognitive test performance and participant enjoyment. PeerJ 2016;4:e2184 [FREE Full text] [doi: 10.7717/peerj.2184] [Medline: 27441120]
32. Miranda AT, Palmer EM. Intrinsic motivation and attentional capture from gamelike features in a visual search task. Behav Res Methods 2014 Mar;46(1):159-172. [doi: 10.3758/s13428-013-0357-7] [Medline: 23835649]
33. Birk MV, Mandryk RL, Bowey J, Buttlar B. The Effects of Adding Premise and Backstory to Psychological Tasks. In: Conference on Human Factors in Computing Systems. 2015 Presented at: CHI'15; April 18-23, 2015; Seoul, Korea.

34. Mekler ED, Brühlmann F, Tuch AN, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Comput Hum Behav 2017 Jun;71:525-534. [doi: 10.1016/j.chb.2015.08.048]
35. Boyle EA, Connolly TM, Hainey T, Boyle JM. Engagement in digital entertainment games: a systematic review. Comput Hum Behav 2012 May;28(3):771-780. [doi: 10.1016/j.chb.2011.11.020]
36. Csikszentmihalyi M, Rathunde K. The measurement of flow in everyday life: toward a theory of emergent motivation. Nebr Symp Motiv 1992;40:57-97. [Medline: 1340523]
37. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
38. Elliot AJ, Dweck CS. Competence and motivation: competence as the core of achievement motivation. In: Handbook of Competence and Motivation. New York, USA: Guilford Publications; 2005.
39. Moneta GB. On the measurement and conceptualization of flow. In: Advances in Flow Research. New York, USA: Springer; 2012.
40. Csikszentmihalyi M. The flow experience and its significance for human psychology. In: Optimal Experience: Psychological Studies of Flow in Consciousness. Cambridge, UK: Cambridge University Press; 1988.
41. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Comput Entertain 2005 Jul 1;3(3):3. [doi: 10.1145/1077246.1077253]
42. Brown E, Cairns A. A Grounded Investigation of Game Immersion. In: CHI '04 Extended Abstracts on Human Factors in Computing Systems. 2004 Presented at: CHI EA'04; April 24-29, 2004; Vienna, Austria. [doi: 10.1145/985921.986048]
43. Bowey JT, Friehs MA, Mandryk RL. Red or Blue Pill: Fostering Identification and Transportation Through Dialogue Choices in RPGs. In: Proceedings of the 14th International Conference on the Foundations of Digital Games. 2019 Presented at: FDG'19; August 26-30, 2019; San Luis Obispo, California. [doi: 10.1145/3337722.3337734]
44. Chen J. Flow in games (and everything else). Commun ACM 2007 Apr 1;50(4):31. [doi: 10.1145/1232743.1232769]
45. Jin SA. 'Toward integrative models of flow': effects of performance, skill, challenge, playfulness, and presence on flow in video games. J Broadcast Electron Media 2012;56(2):169-186 [FREE Full text]
46. Alexandrovsky D, Friehs A, Birk MV, Yates RK, Mandryk RL. Game Dynamics that Support Snacking, not Feasting. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2019 Presented at: CHI PLAY'19; October 22-25, 2019; Barcelona, Spain. [doi: 10.1145/3311350.3347151]
47. Deci EL, Ryan RM. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
48. Ryan R, Deci E. Self-determination theory: an organismic dialectical perspective. In: The Handbook of Self-Determination Research. New York, USA: Springer; 2002.
49. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
50. Peng W, Lin J, Pfeiffer KA, Winn B. Need satisfaction supportive game features as motivational determinants: an experimental study of a self-determination theory guided exergame. Media Psychol 2012 May 18;15(2):175-196. [doi: 10.1080/15213269.2012.673850]
51. Birk MV, Atkins C, Bowey JT, Mandryk RL. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2858036.2858062]
52. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol 2013;105(2):249-265. [doi: 10.1037/a0031311]
53. Birk MV, Friehs MA, Mandryk RL. Age-Based Preferences and Player Experience: A Crowdsourced Cross-sectional Study. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY'17; October 15-18, 2017; Amsterdam, The Netherlands. [doi: 10.1145/3116595.3116608]
54. Nicholson S. A recipe for meaningful gamification. In: Gamification in Education and Business. New York, USA: Springer; 2015.
55. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
56. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
57. Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: a model and a method. J Exp Psychol Hum Percept Perform 1984 Apr;10(2):276-291. [doi: 10.1037//0096-1523.10.2.276] [Medline: 6232345]
58. Lappin JS, Eriksen CW. Use of a delayed signal to stop a visual reaction-time response. J. Exp. Psychol 1966;72(6):805-811 [FREE Full text]
59. Jennings JR, van der Molen MW, Pelham W, Debski KB, Hoza B. Inhibition in boys with attention deficit hyperactivity disorder as indexed by heart rate change. Dev Psychol 1997 Mar;33(2):308-318. [doi: 10.1037/0012-1649.33.2.308] [Medline: 9147839]

60. Schachar R, Logan GD. Impulsivity and inhibitory control in normal development and childhood psychopathology. Dev Psychol 1990 Sep;26(5):710-720. [doi: 10.1037/0012-1649.26.5.710]
61. Hoptman MJ, Ardekani BA, Butler PD, Nierenberg J, Javitt DC, Lim KO. DTI and impulsivity in schizophrenia: a first voxelwise correlational analysis. Neuroreport 2004 Nov 15;15(16):2467-2470 [FREE Full text] [doi: 10.1097/00001756-200411150-00007] [Medline: 15538176]
62. Kiehl KA, Smith AM, Hare RD, Liddle PF. An event-related potential investigation of response inhibition in schizophrenia and psychopathy. Biol Psychiatry 2000 Aug 1;48(3):210-221. [doi: 10.1016/s0006-3223(00)00834-9] [Medline: 10924664]
63. Wang C, Chang C, Liang Y, Shih C, Chiu W, Tseng P, et al. Open vs. closed skill sports and the modulation of inhibitory control. PLoS One 2013;8(2):e55773 [FREE Full text] [doi: 10.1371/journal.pone.0055773] [Medline: 23418458]
64. Tsai C. The effectiveness of exercise intervention on inhibitory control in children with developmental coordination disorder: using a visuospatial attention paradigm as a model. Res Dev Disabil 2009;30(6):1268-1280. [doi: 10.1016/j.ridd.2009.05.001] [Medline: 19497707]
65. Friehs MA, Frings C. Pimping inhibition: Anodal tDCS enhances stop-signal reaction time. J Exp Psychol Hum Percept Perform 2018 Dec;44(12):1933-1945. [doi: 10.1037/xhp0000579] [Medline: 30299126]
66. Friehs MA, Frings C. Cathodal tDCS increases stop-signal reaction time. Cogn Affect Behav Neurosci 2019 Oct;19(5):1129-1142. [doi: 10.3758/s13415-019-00740-0] [Medline: 31313248]
67. Soreni N, Crosbie J, Ickowicz A, Schachar R. Stop signal and Conners' continuous performance tasks: test--retest reliability of two inhibition measures in ADHD children. J Atten Disord 2009 Sep;13(2):137-143. [doi: 10.1177/1087054708326110] [Medline: 19429883]
68. Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? J Abnorm Psychol 2005 May;114(2):216-222. [doi: 10.1037/0021-843X.114.2.216] [Medline: 15869352]
69. Lipszyc J, Schachar R. Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. J Int Neuropsychol Soc 2010 Nov;16(6):1064-1076. [doi: 10.1017/S1355617710000895] [Medline: 20719043]
70. Delisle J, Braun CM. A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type). Arch Clin Neuropsychol 2011 Nov;26(7):602-613. [doi: 10.1093/arclin/acr043] [Medline: 21653627]
71. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18(6):643-662. [doi: 10.1037/h0054651]
72. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 1974 Jan;16(1):143-149. [doi: 10.3758/bf03203267]
73. Güldenpenning I, Steinke A, Koester D, Schack T. Athletes and novices are differently capable to recognize feint and non-feint actions. Exp Brain Res 2013 Oct;230(3):333-343. [doi: 10.1007/s00221-013-3658-2] [Medline: 23955103]
74. Kunde W, Skirde S, Weigelt M. Trust my face: cognitive factors of head fakes in sports. J Exp Psychol Appl 2011 Jun;17(2):110-127. [doi: 10.1037/a0023756] [Medline: 21604910]
75. Güldenpenning I, Schütz C, Weigelt M, Kunde W. Is the head-fake effect in basketball robust against practice? Analyses of trial-by-trial adaptations, frequency distributions, and mixture effects to evaluate effects of practice. Psychol Res 2020 Apr;84(3):823-833. [doi: 10.1007/s00426-018-1078-4] [Medline: 30128660]
76. Frings C, Brinkmann T, Friehs MA, van Lipzig T. Single session tDCS over the left DLPFC disrupts interference processing. Brain Cogn 2018 Feb;120:1-7. [doi: 10.1016/j.bandc.2017.11.005] [Medline: 29202318]
77. Friehs MA, Güldenpenning I, Frings C, Weigelt M. Electrify your game! Anodal tdcs increases the resistance to head fakes in basketball. J Cogn Enhanc 2019 Apr 24;4(1):62-70. [doi: 10.1007/s41465-019-00133-8]
78. Blascovich J, Loomis J, Beall AC, Swinth KR, Hoyt CL, Bailenson JN. Authors' response: immersive virtual environment technology: just another methodological tool for social psychology? Psychol Inq 2002 Apr;13(2):146-149. [doi: 10.1207/s15327965pli1302_03]
79. McDermott R. Experimental methodology in political science. Polit Anal 2017 Jan 4;10(4):325-342. [doi: 10.1093/pan/10.4.325]
80. Highhouse S. Designing experiments that generalize. Organ Res Methods 2007 Jul 23;12(3):554-566. [doi: 10.1177/1094428107300396]
81. Dobbins GH, Lane IM, Steiner DD. A note on the role of laboratory methodologies in applied behavioural research: don't throw out the baby with the bath water. J Organ Behav 1988;9(3):281-286 [FREE Full text]
82. Jackson SA, Marsh W. Development and validation of a scale to measure optimal experience: the flow state scale. J Sport Exerc Psychol 1996;18(1):17-35 [FREE Full text] [doi: 10.1123/jsep.18.1.17]
83. Verbruggen F, Stevens T, Chambers CD. Proactive and reactive stopping when distracted: an attentional account. J Exp Psychol Hum Percept Perform 2014 Aug;40(4):1295-1300 [FREE Full text] [doi: 10.1037/a0036542] [Medline: 24842070]
84. Faul F, ErdFelder E, Lang AG, Buchner A. G*power 3.1 manual. Behav Res Methods 2007;39(2):175-191 [FREE Full text]
85. Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife 2019 Apr 29;8:- [FREE Full text] [doi: 10.7554/eLife.46323] [Medline: 31033438]

86. Verbruggen F, Logan GD. Evidence for capacity sharing when stopping. Cognition 2015 Sep;142:81-95 [FREE Full text] [doi: 10.1016/j.cognition.2015.05.014] [Medline: 26036922]
87. David FN, Tukey JW. Exploratory data analysis. Biometrics 1977 Dec;33(4):768. [doi: 10.2307/2529486]
88. Verbruggen F, Chambers CD, Logan GD. Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. Psychol Sci 2013 Mar 1;24(3):352-362 [FREE Full text] [doi: 10.1177/0956797612457390] [Medline: 23399493]
89. Rouder JN, Morey RD, Verhagen J, Province JM, Wagenmakers E. Is There a Free Lunch in Inference? Top Cogn Sci 2016 Jul;8(3):520-547 [FREE Full text] [doi: 10.1111/tops.12214] [Medline: 27489199]
90. Wagenmakers E, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon Bull Rev 2018 Feb;25(1):35-57 [FREE Full text] [doi: 10.3758/s13423-017-1343-3] [Medline: 28779455]
91. Wagenmakers E, Wetzels R, Borsboom D, van der Maas HL. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). J Pers Soc Psychol 2011 Mar;100(3):426-432. [doi: 10.1037/a0022790] [Medline: 21280965]
92. Wagenmakers E, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: example applications with JASP. Psychon Bull Rev 2018 Feb;25(1):58-76 [FREE Full text] [doi: 10.3758/s13423-017-1323-7] [Medline: 28685272]
93. Stigler SM. Regression towards the mean, historically considered. Stat Methods Med Res 1997 Jun;6(2):103-114. [doi: 10.1177/096228029700600202] [Medline: 9261910]
94. Bland JM, Altman DG. Regression towards the mean. Br Med J 1994 Jun 4;308(6942):1499 [FREE Full text] [doi: 10.1136/bmj.308.6942.1499] [Medline: 8019287]
95. Logan GD. The point of no return: a fundamental limit on the ability to control thought and action. Q J Exp Psychol (Hove) 2015;68(5):833-857 [FREE Full text] [doi: 10.1080/17470218.2015.1008020] [Medline: 25633089]
96. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. Trends Cogn Sci 2004 Apr;8(4):170-177. [doi: 10.1016/j.tics.2004.02.010] [Medline: 15050513]
97. Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. J Neurosci 2007 Apr 4;27(14):3743-3752 [FREE Full text] [doi: 10.1523/JNEUROSCI.0519-07.2007] [Medline: 17409238]
98. Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, et al. Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. Neuroimage 2012 Feb 1;59(3):2860-2870 [FREE Full text] [doi: 10.1016/j.neuroimage.2011.09.049] [Medline: 21979383]
99. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn Sci 2014 Apr;18(4):177-185. [doi: 10.1016/j.tics.2013.12.003] [Medline: 24440116]
100. Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. Neuroimage 2002 Dec;17(4):1820-1829. [doi: 10.1006/nimg.2002.1326] [Medline: 12498755]
101. Wang M, Dong G, Wang L, Zheng H, Potenza MN. Brain responses during strategic online gaming of varying proficiencies: implications for better gaming. Brain Behav 2018 Aug;8(8):e01076 [FREE Full text] [doi: 10.1002/brb3.1076] [Medline: 30020566]
102. Catarino A, Küpper CS, Werner-Seidler A, Dalgleish T, Anderson MC. Failing to forget: inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. Psychol Sci 2015 May;26(5):604-616 [FREE Full text] [doi: 10.1177/0956797615569889] [Medline: 25847536]
103. Falconer E, Bryant R, Felmingham KL, Kemp AH, Gordon E, Peduto A, et al. The neural networks of inhibitory control in posttraumatic stress disorder. J Psychiatry Neurosci 2008 Sep;33(5):413-422 [FREE Full text] [Medline: 18787658]
104. Birk MV, Mandryk RL. The Benefits of Digital Games for the Assessment and Treatment of Mental Health. In: Computing and Mental Health Workshop. 2016 Presented at: CHI'16; May 8, 2016; San Jose, CA.
105. Birk MV, Wadley G, Abeele VV, Mandryk R, Torous J. Video games for mental health. Interactions 2019 Jun 26;26(4):32-36. [doi: 10.1145/3328483]
106. Hedge C, Powell G, Sumner P. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. Behav Res Methods 2018 Jun;50(3):1166-1186 [FREE Full text] [doi: 10.3758/s13428-017-0935-1] [Medline: 28726177]
107. Hooshyar D, Ahmad RB, Yousefi M, Fathi M, Horng S, Lim H. Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. Comput Educ 2016 Mar;94:18-36. [doi: 10.1016/j.compedu.2015.10.013]
108. Bevilacqua F, Engström H, Backlund P. Game-calibrated and user-tailored remote detection of stress and boredom in games. Sensors (Basel) 2019 Jun 28;19(13):- [FREE Full text] [doi: 10.3390/s19132877] [Medline: 31261716]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ANOVA: analysis of variance
AOI: area of interest
BF: Bayes factor
FSS: Flow State Scale
IMI: Intrinsic Motivation Inventory
MANOVA: multivariate analysis of variance
OCD: obsessive-compulsive disorder
RT: reaction time
SDT: Self-Determination Theory
SSD: stop-signal delay
SSG: stop-signal game
SSRT: stop-signal reaction time
SST: stop-signal task

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by A De Marchi, N Khalili-Mahani, H Söbke; comments to author 12.06.20; revised version received 16.06.20; accepted 25.06.20; published 08.09.20.

Please cite as:
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;8(3):e17810
URL: https://games.jmir.org/2020/3/e17810
doi:10.2196/17810
PMID:32897233

Original Paper

Effects of Avatar Perspective on Joint Excursions Used to Play Virtual Dodgeball: Within-Subject Comparative Study

Susanne M van der Veen[1], PhD; Alexander Stamenkovic[1], PhD; Megan E Applegate[2], PhD; Samuel T Leitkam[3], PhD; Christopher R France[4], PhD; James S Thomas[1], PhD

[1]Virginia Commonwealth University, Richmond, VA, United States
[2]Harvard University, Boston, MA, United States
[3]McGill University, Montreal, QC, Canada
[4]Ohio University, Athens, OH, United States

Corresponding Author:
Susanne M van der Veen, PhD
Virginia Commonwealth University
900 E Leigh St
4th Floor
Richmond, VA, 23298
United States
Phone: 1 7405416324
Email: smvanderveen@vcu.edu

Abstract

Background: Visual representation of oneself is likely to affect movement patterns. Prior work in virtual dodgeball showed greater excursion of the ankles, knees, hips, spine, and shoulder occurs when presented in the first-person perspective compared to the third-person perspective. However, the mode of presentation differed between the two conditions such that a head-mounted display was used to present the avatar in the first-person perspective, but a 3D television (3DTV) display was used to present the avatar in the third-person. Thus, it is unknown whether changes in joint excursions are driven by the visual display (head-mounted display versus 3DTV) or avatar perspective during virtual gameplay.

Objective: This study aimed to determine the influence of avatar perspective on joint excursion in healthy individuals playing virtual dodgeball using a head-mounted display.

Methods: Participants (n=29, 15 male, 14 female) performed full-body movements to intercept launched virtual targets presented in a game of virtual dodgeball using a head-mounted display. Two avatar perspectives were compared during each session of gameplay. A first-person perspective was created by placing the center of the displayed content at the bridge of the participant's nose, while a third-person perspective was created by placing the camera view at the participant's eye level but set 1 m behind the participant avatar. During gameplay, virtual dodgeballs were launched at a consistent velocity of 30 m/s to one of nine locations determined by a combination of three different intended impact heights and three different directions (left, center, or right) based on subject anthropometrics. Joint kinematics and angular excursions of the ankles, knees, hips, lumbar spine, elbows, and shoulders were assessed.

Results: The change in joint excursions from initial posture to the interception of the virtual dodgeball were averaged across trials. Separate repeated-measures ANOVAs revealed greater excursions of the ankle (P=.010), knee (P=.001), hip (P=.0014), spine (P=.001), and shoulder (P=.001) joints while playing virtual dodgeball in the first versus third-person perspective. Aligning with the expectations, there was a significant effect of impact height on joint excursions.

Conclusions: As clinicians develop treatment strategies in virtual reality to shape motion in orthopedic populations, it is important to be aware that changes in avatar perspective can significantly influence motor behavior. These data are important for the development of virtual reality assessment and treatment tools that are becoming increasingly practical for home and clinic-based rehabilitation.

(JMIR Serious Games 2020;8(3):e18888) doi:10.2196/18888

KEYWORDS

virtual reality; avatar perspective; reaching; joint excursion; exergaming; exercise rehabilitation; head mounted display

Introduction

Virtual reality (VR)-based interventions hold great potential for rehabilitation, as they can be used to both assess motor coordination and elicit specific movements in a manner that is simultaneously engaging and therapeutic. Further, visual stimuli can be easily presented in VR and manipulated in real-time, providing insight into the neural mechanisms underpinning sensorimotor control of movement. While VR systems are becoming more affordable and readily available for in-home rehabilitation applications, the effects of different avatar perspectives on motor behavior are poorly understood and need to be studied to optimize interventions. For example, the use of home devices such as the Kinect sensor, which tracks and presents an avatar in a third-person perspective, may result in very different motor behavior when compared to the same tasks being presented from the first-person perspective.

Virtual reality has been used to shape motion in orthopedic [1-3] and neurologic patient populations [4-6], with reports showing significant effects on pain relief, joint mobility and motor function [7]. However, the vast differences in methodology, especially concerning visual display type, avatar perspective, and level of gameplay immersion, make it difficult to draw broad conclusions about which features are driving the efficacy of VR treatments [7]. Visual environments in VR can present 3-dimensional (3D) images across a variety of display devices, including head mounted-displays (HMD) [6] and 3D televisions (3DTV) [8,9]. The different methods used to present visual scenes can affect how the virtual environment is perceived and thus can influence not only motor behavior [8] but pain responses as well [10].

While Thomas and colleagues have shown that avatar perspective influences joint excursions in full-body reaching tasks [9] as well as during VR gameplay [8], it is unknown if the differences in joint excursions were driven by avatar perspective (first- or third-person) or by display type (HMD or 3DTV). Ustinova et al [11] reported that individuals reached further and preferred tasks when the camera perspective was oriented at angles from 45 to 77.5 degrees relative to the location of the avatar as compared to the camera oriented at zero degrees (ie, directly behind the avatar) [11]. The increased segment displacement was accompanied by a slightly larger displacement of the whole-body center of mass, yet these reaches were always made to varying degrees in the third-person avatar perspective.

To date, no study has compared the effect of the avatar perspective alone on the apportionment of joint excursions in VR tasks while keeping the display type constant. This study was designed to determine the effects of avatar perspective (ie, first- versus third-person) on joint excursions of healthy participants engaged in full-body movements during a VR dodgeball game presented on an HMD. Based on existing studies [8,9,11], we predicted that joint excursions would be greater in the first person versus third-person perspective.

Methods

Recruitment

We recruited 29 healthy young adults (15 male, 14 female) aged 18-35 years (mean ± SD, 23 ± 1.62 years, range 20-28 years). Exclusion criteria included a history of a low back injury, low back pain within the last 6 months, and any orthopedic, neurological, or visual impairment that would prevent participation. This study was approved by the Institutional Review Board of Ohio University, and written informed consent was obtained at the beginning of the session.

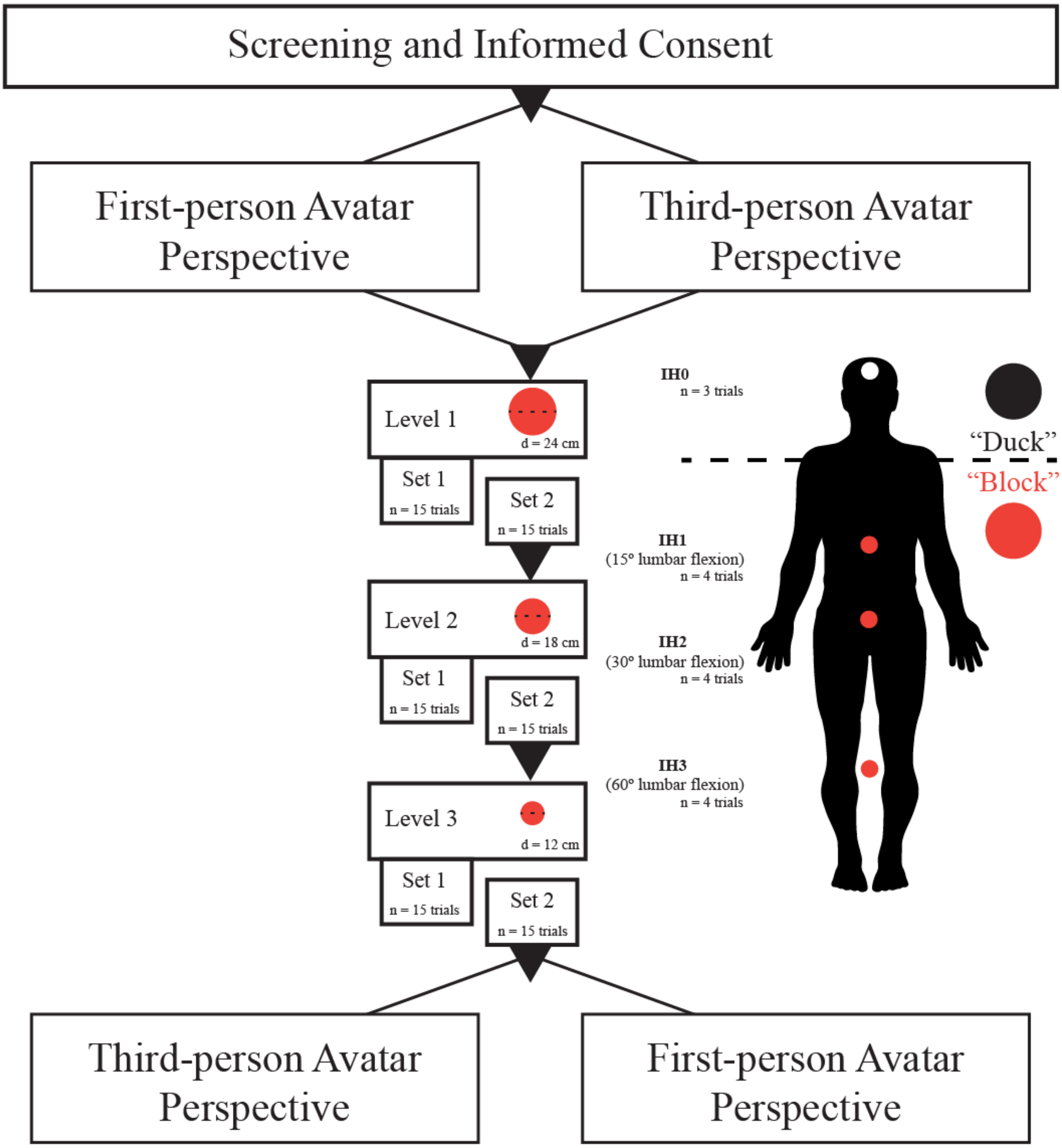
Instrumentation

Movement of light-reflective marker clusters attached to the head, upper arms, forearms, hands, trunk, pelvis, thighs, shanks, and feet was tracked using a 10-camera Vicon Bonita system sampling at 100 Hz. This optoelectric-based kinematic system can track the 3D coordinates of light reflective marker clusters attached to the participant with a spatial resolution of 0.1 mm. The time-series joint angle data were derived from the 3D segment coordinate data using an Euler angle sequence of (1) flexion-extension, (2) lateral bending, and (3) axial rotation using MotionMonitor software (The MotionMonitor) [12]. Joint excursions were defined as the change in joint angle from initial standing posture to posture at interception of launched virtual balls.

Procedures

The study employed a within-subjects design, in which participation consisted of gameplay during two virtual sessions of dodgeball. Each session involved gameplay presented in either the first- or third-person perspective. A first-person perspective was created by placing the center of the displayed content at the bridge of the participant’s nose, while a third-person perspective was created by placing the camera view at the participant’s eye level and set 1 m behind the participant avatar. The order of avatar perspective was randomized and counterbalanced such that half the study cohort began with gameplay using a first-person perspective, and half began with a third-person perspective (see Figure 1). During gameplay, participants competed against four virtual opponents, and the object was to block or avoid virtual balls launched by the four opponents. The intended impact heights of the launched virtual balls were identical between the two avatar perspectives (see Gameplay section for a description of impact height determination). Participants earned game points (which were associated with actual cash rewards) by either successfully blocking the virtual balls (see Figure 1, "Duck") or, in the case of certain colored balls, avoiding contact with them.

Figure 1. A flow chart of the recruitment, randomization, and dodgeball gameplay. The red dots representing impact heights are solely for visualization.



Virtual Environment

Vizard software (WorldViz) was used to develop the virtual environment and control all presented graphics and audio stimuli, including the opposing team's avatars. The six degrees of freedom kinematic data from the clusters of light reflective markers placed on the participant was streamed to the game environment at 100 Hz using Vicon Tracker software, allowing for near real-time presentation of the participant's avatar (39 ms latency). The MotionMonitor software was used to control bidirectional communication with Vizard, set game parameters and target locations, and record all kinematic data during the experimental testing session. Participants viewed their avatar from a first- or third-person perspective via an HMD (Oculus Rift Developers Kit 2, Oculus). The HMD display provided a 90° horizontal field of view, and the refresh rate was fixed at 75 Hz/eye.

Gameplay

The game environment was an indoor basketball arena, with the participant positioned at the free-throw line on one side of the court, facing the four virtual opponents positioned on the opposite free-throw line. The opposing players moved 3 m fore-aft and 3m left-right in random order. Virtual balls were launched with a speed of 30 m/s every 3.3 ± 0.3 seconds in a randomized order from each of the four virtual opponents. To indicate the launching of a virtual ball, an opponent would flash either green or yellow 300 ms before release. The color signified whether participants had to block the oncoming ball (green flash, red ball), using the virtual ball co-located with a physical dodgeball in the real world (and held between the participant's hands), or avoid the ball by "ducking" below it (yellow flash, black ball). This second condition was accompanied by an audible duck quacking sound to emphasize the goal of ducking to avoid the ball. A large scoreboard was positioned at the opposite end of the arena (above the opponents) so that participants could track their performance and cash rewards earned. Additional sound effects included crowd cheering, buzzers, and referee whistles. An instrumented participant engaged in virtual dodgeball with the HMD is shown in Figure 2. A gameplay session consisted of three levels lasting approximately 2 minutes. Each game level consisted of 2 sets with 15 balls in each set. The intended impact locations of 12 of the 15 balls launched were distributed amongst the three impact heights normalized to the participant's arm length, trunk length, and hip height [8,9,13-16]. For example, during gameplay, the participant could successfully block a virtual ball launched to impact height 1 (ie, IH1, the highest impact height) simply by flexing at the lumbar spine 15° with the elbow fully extended and the shoulder flexed to 90°. In contrast, 60° of lumbar flexion would be required to intercept a virtual ball under the same conditions launched to IH3 (ie, the lowest impact height, see Figure 3). Three balls were launched at each impact height, one to intersect the participant at their midline and one each 20 cm to the left and right of the midline. An additional three balls were launched targeting the head of the avatar to elicit avoidance of the ball by ducking. As balls that required ducking elicited different movement behaviors (ie, object avoidance rather than interception via a reaching motion), these were not investigated in the current study. It is important to note that the order in which each virtual ball was launched at different impact heights was permutated at each round of gameplay. Each level saw a reduction in the size of the dodgeball thrown, starting with a standard-sized dodgeball (diameter: 24 cm) in level 1, reducing across levels (diameter: 18 cm, level 2; diameter: 12 cm, level 3) in order to increase gameplay challenge (see Figure 1). Gameplay performance was updated in real-time and displayed on the virtual scoreboard, with the participant earning progressively higher rewards for each successful block at each level of play (Practice Level = 1¢, Level 1 = 2¢, Level 2 = 5¢, Level 3 = 10¢). Similar to the popular game of dodgeball, success was determined when the oncoming dodgeball was intercepted before contact with the virtual avatar was made.

Figure 2. Participant instrumented and engaged in virtual dodgeball using a head-mounted display. Differences in the representation of the avatar in the virtual environment (upper panels) and during gameplay (bottom panels) are shown for both the first-person (left panels) and third-person perspectives (right panels).

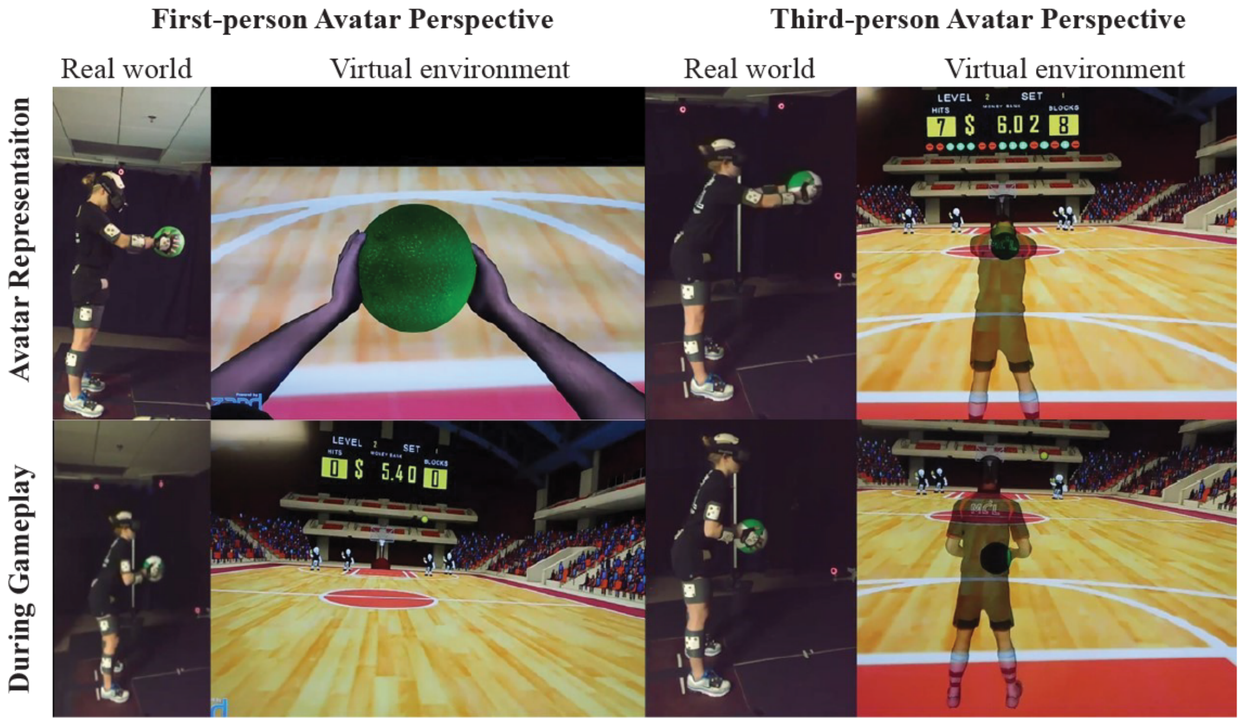
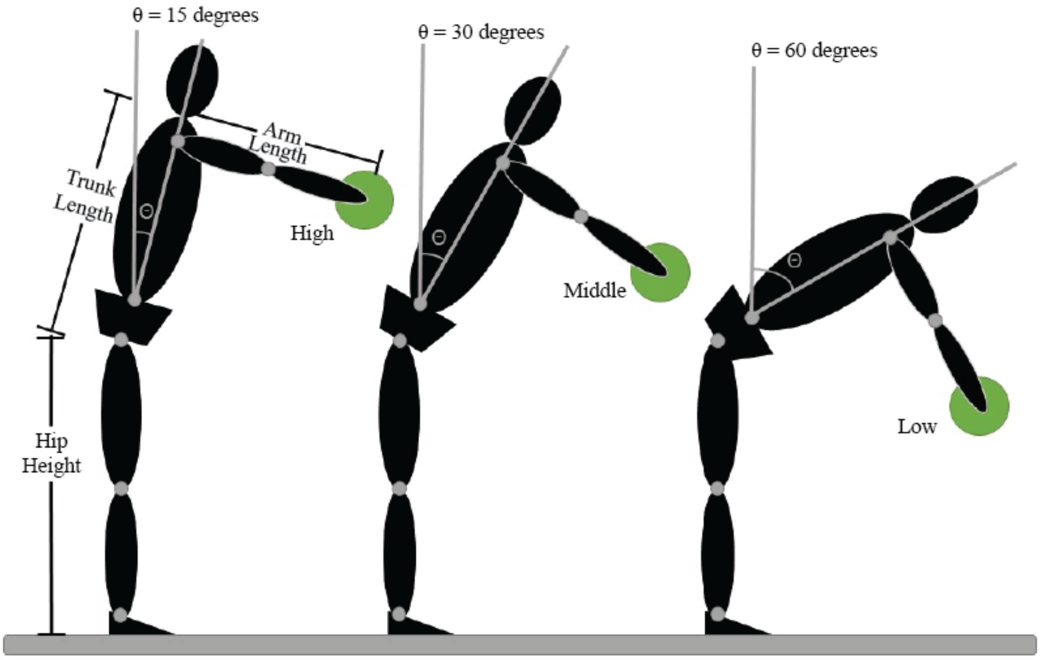


Figure 3. Methods for computing location of the impact heights (IH1: highest; IH3: lowest) of the launched virtual balls for a single game level. Impact heights were tailored for each participant based on their hip height, trunk length, and arm length to varying degrees of lumbar spine flexion.



Conversely, the participant lost cash rewards for each failure to block an oncoming ball from hitting their virtual avatar. Each player started the game with a cash balance on the scoreboard such that if they failed on every launched or presented ball, their cash balance would be zero. The average gameplay session lasted approximately 15 minutes. Following each session, participants rated their overall effort using the NASA Task Load Index (TLX). The NASA TLX is a multidimensional assessment that rates perceived workload to assess system performance [17]. Specifically, participants provide experience ratings of 1 (very low) to 7 (very high) along six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. In the context of assessing the avatar perspective on game success, this measure provides insight into differences in the perceived workload performing nearly identical tasks.

Data Reduction and Analysis

As gameplay occurred with the dodgeball centered between the left and right hands of the participants, and initial examination of joint excursions was nearly identical for the left and right. Therefore, analyses have been limited to comparisons for the right side only. First, the time-series position vector of the right index fingertip was smoothed using a 41-point fourth-order Savitzky-Golay filter [18]. That is, at each time sample, fourth-order polynomials were fit in the least-squares sense to the data at that point and 20 neighboring samples on each side. The polynomial coefficients were then used to determine velocity. Movement onset was determined from a backward search from peak velocity and defined as the point where velocity was ≤5% peak velocity. Target contact was defined as the point where velocity was ≤5% peak velocity using a forward search from peak velocity. Movement time was determined from movement onset to target contact, as specified above. The change in joint angles (ie, ankle, knee, hip, spine, shoulder, and elbow) and displacement of the whole-body center of mass, based on the combination of segment center of mass [19], along the antero-posterior, mediolateral, and vertical axes were calculated from movement onset to target contact. To determine hand position at target contact, we first calculated the centroid of the hands from the x, y, and z position traces from marker clusters on the left and right hands and adjusted this to the centroid of the left and right ankle joint. We then determined the hand position at target contact for the antero-posterior, mediolateral, and vertical axes.

Statistical Analysis

An initial power calculation using G*Power 3.19 [28] based on joint excursions observed during pilot testing found that 14 participants were necessary to determine the within-subject effects of the avatar perspective with 80% power, assuming alpha = .05, a correlation between measures of 0 .5, and an effect size of f = 0.4 (large effect). As we were primarily interested in the influence of the avatar perspective on movement behavior, we only analyzed trials using the standard dodgeball size. Separate 3-way repeated measures analyses of variance were performed for each dependent measure, with sex as the between-subjects variable, and perspective (first person, third person) and impact height (IH1-IH3) as within-subject variables. Dependent measures included: (1) movement time, (2) hand position at target contact in the antero-posterior, mediolateral, and vertical planes, (3) joint angular excursions of the right ankle, knee, hip, spine, shoulder, and elbow, (4) displacement of the center of mass (ie, antero-posterior, mediolateral, vertical), and (5) success rate for the standard dodgeball size. Post-hoc analyses were performed using the method of least significant differences. Interactions were examined using a simple effects model. The NASA TLX data were analyzed using paired t-tests with Bonferroni correction for multiple comparisons. All statistical analyses were completed in SPSS 22 (IBM).

Results

Movement Time

There was no effect of avatar perspective on movement time, but movement time did differ as a function of impact height ($F_{2,27}$=9.181, P<.001). Specifically, movement time was less for interception of virtual dodgeballs launched to the highest (574 ms ± 40 ms) versus the middle (648 ms ± 27 ms, P=.006) and low impact heights (671 ms ± 23 ms, P=.002). There were no interactions of perspective by impact height on movement time.

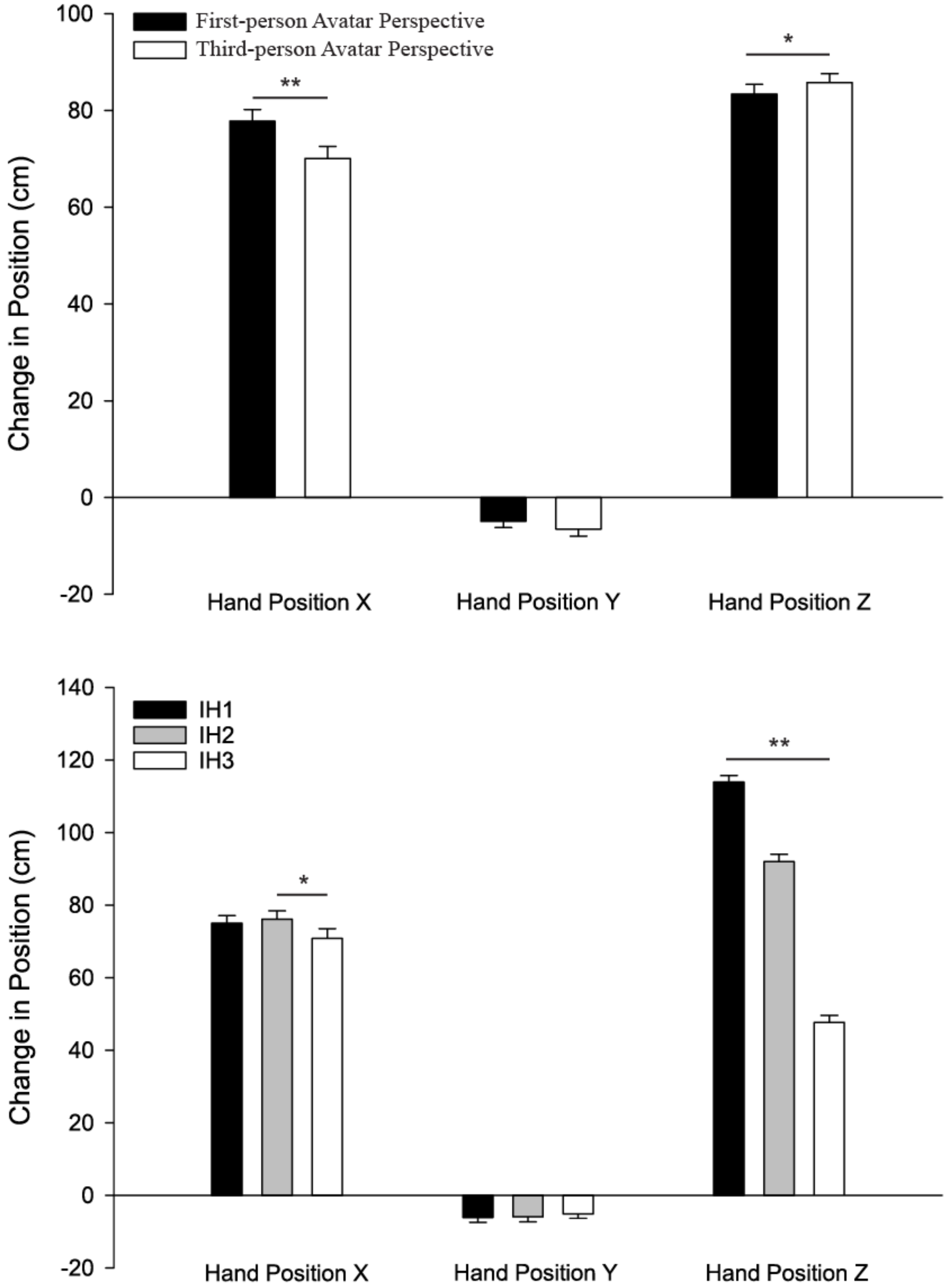
Hand Position at Ball Contact

Antero-posterior hand position at ball contact was further forward when the participant's avatar was presented in a first-person (76.7 cm ± 2.2 cm) versus a third-person (69.1 cm ± 2.7 cm) perspective ($F_{1,27}$=20.410, P<.001). Antero-posterior hand position at ball contact was also influenced by impact height ($F_{2,26}$=12.980, P<.001). Specifically, participants did not reach as far forward for virtual balls launched to the lowest impact height (68.1 cm ± 2.6 cm) as compared to the middle impact height (75.7 cm ± 2.4 cm, P=.010) and the highest impact height (75.0 cm ± 2.2 cm, P=.013). There were no interaction effects of perspective and impact height.

There was an interaction of avatar perspective and impact height on vertical hand position at ball contact ($F_{2,26}$= 4.237, P=.020, see Figure 4). Follow-up analyses revealed that the avatar perspective resulted in differences in vertical hand position at ball contact for the lowest impact height (IH3) (first-person 45.0 cm ± 2.4 cm and third-person 50.4 cm ± 2.1 cm, P=.023). Post-hoc analyses revealed that participants had to reach lower to intercept virtual balls launched to the lowest impact height (IH3) (47.7 cm ± 1.9 cm) versus the middle (IH2) (92.0 cm ± 2.0 cm) and highest (IH1) (113.9 cm ± 1.8 cm) impact heights (P<.001 for all comparisons). Additionally, an interaction effect of sex and impact height was found ($F_{1,27}$=14.158, P<.001); post-hoc analysis showed males had lower hand position at ball contact than females for IH1 and IH2 (females, IH1: 117.3 cm ± 2.6 cm, IH2: 93.5 cm ± 2.8 cm; males, IH1: 115 cm ± 2.5 cm, IH2 90.6 cm ± 2.8 cm, P=.036 and P=.040 respectively).

No significant differences along the mediolateral axis were found.

Figure 4. The top panel shows the effect of avatar perspective (first- and third-person) on mean hand position averaged across all three impact heights. The bottom panel shows the effect of target impact height (IH1: high, IH2: middle, IH3: low). Errors bars indicate the standard error of the mean, ***P*<.001; **P*<.05).



Joint Excursions

During gameplay presented in the first-person versus third-person perspective, participants increased excursions of the ankle, $F_{1,27}$=7.570, *P*=.010, knee ($F_{1,27}$=12.797, *P*=.001), hip ($F_{1,27}$=6.899, *P*=.014), spine ($F_{1,27}$=14.515, *P*=.001), and shoulder ($F_{1,27}$=13.223, *P*=.001) (see Figures 5 and 6). However, there was no effect of avatar perspective on elbow excursions across all impact heights.

As expected, most joint excursions increased as a function of impact height. Specifically, participants used greater excursions of the ankle ($F_{2,26}$=25.050, p<0.001), knee ($F_{2,26}$=51.198, *P*<.001), hip ($F_{2,26}$=37.538, *P*<.001), spine ($F_{2,26}$=74.462, *P*<.001), and elbow ($F_{2,26}$=3.685, *P*=.039), from the highest to the lowest impact height. However, there was no effect of impact height on shoulder excursions.

An interaction between sex and perspective was found for elbow flexion ($F_{1,27}$=5.468, P=.027), where males showed greater elbow flexion in first-person avatar perspective (11.7° ± 3.0°) than in third-person avatar perspective (10.8° ± 4.7°, P=.005).

Figure 5. Average effects of first-person avatar perspective versus third-person avatar perspective for IH1 (left panel), IH2 (middle panel), and IH3 (right panel) on the posture adopted at target intercept while playing virtual dodgeball using a head-mounted display.

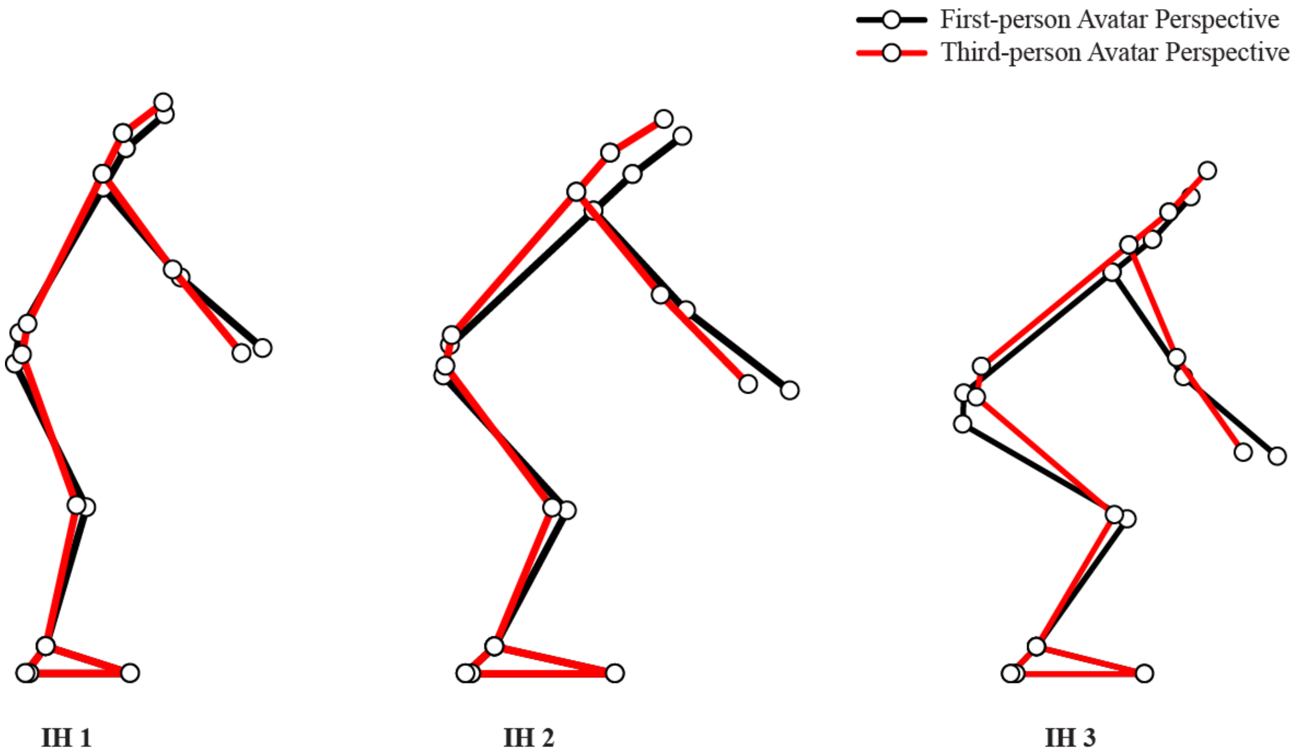
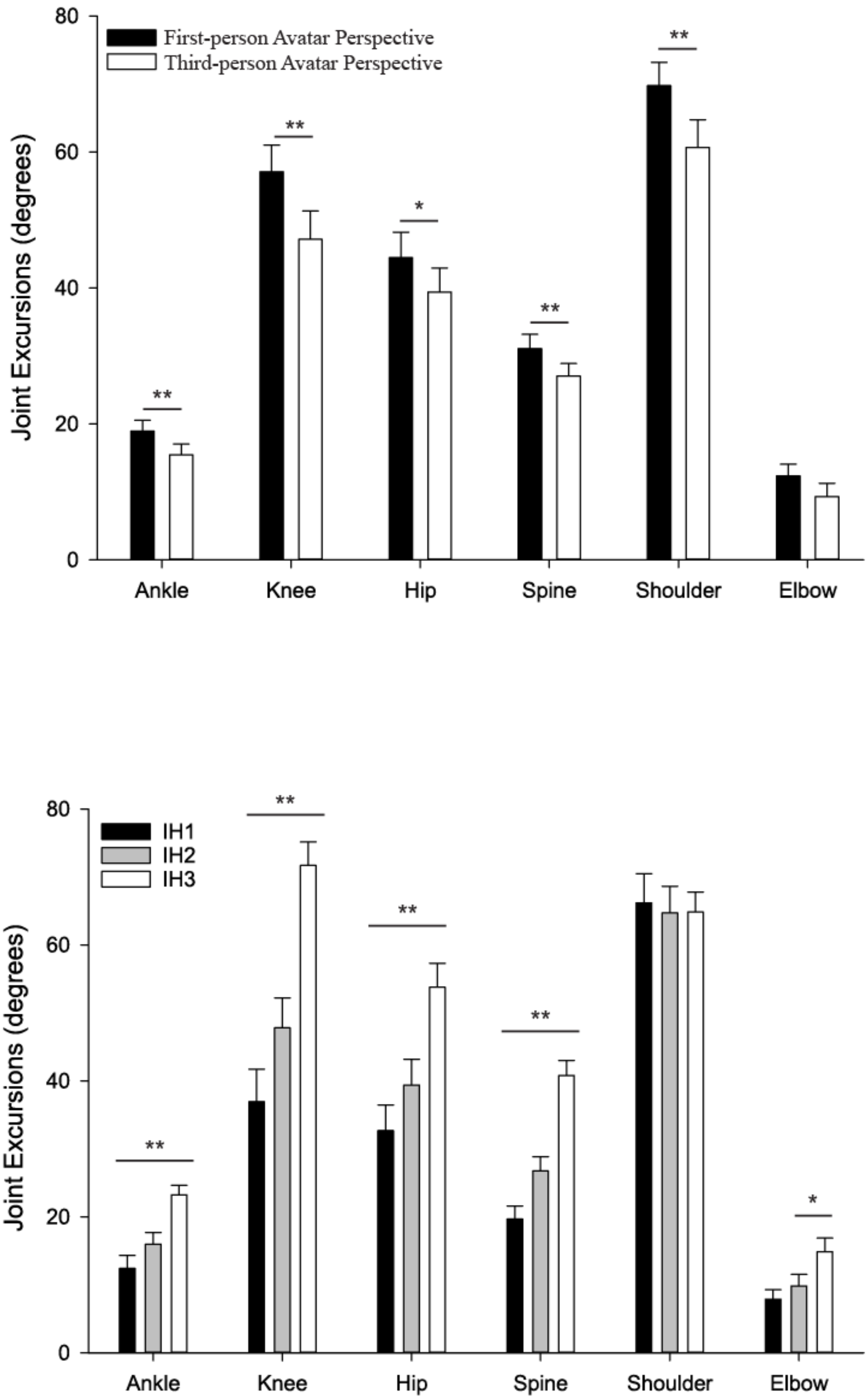


Figure 6. The top panel shows the effect of avatar perspective on mean joint excursions averaged across all three impact heights of the ankle, knee, hip, spine, shoulder, and elbow. The bottom panel shows the effect of target impact height (IH1: high, IH2: middle, IH3: low). Error bars indicate the standard error of the mean, ***P*<.001; **P*<.05.



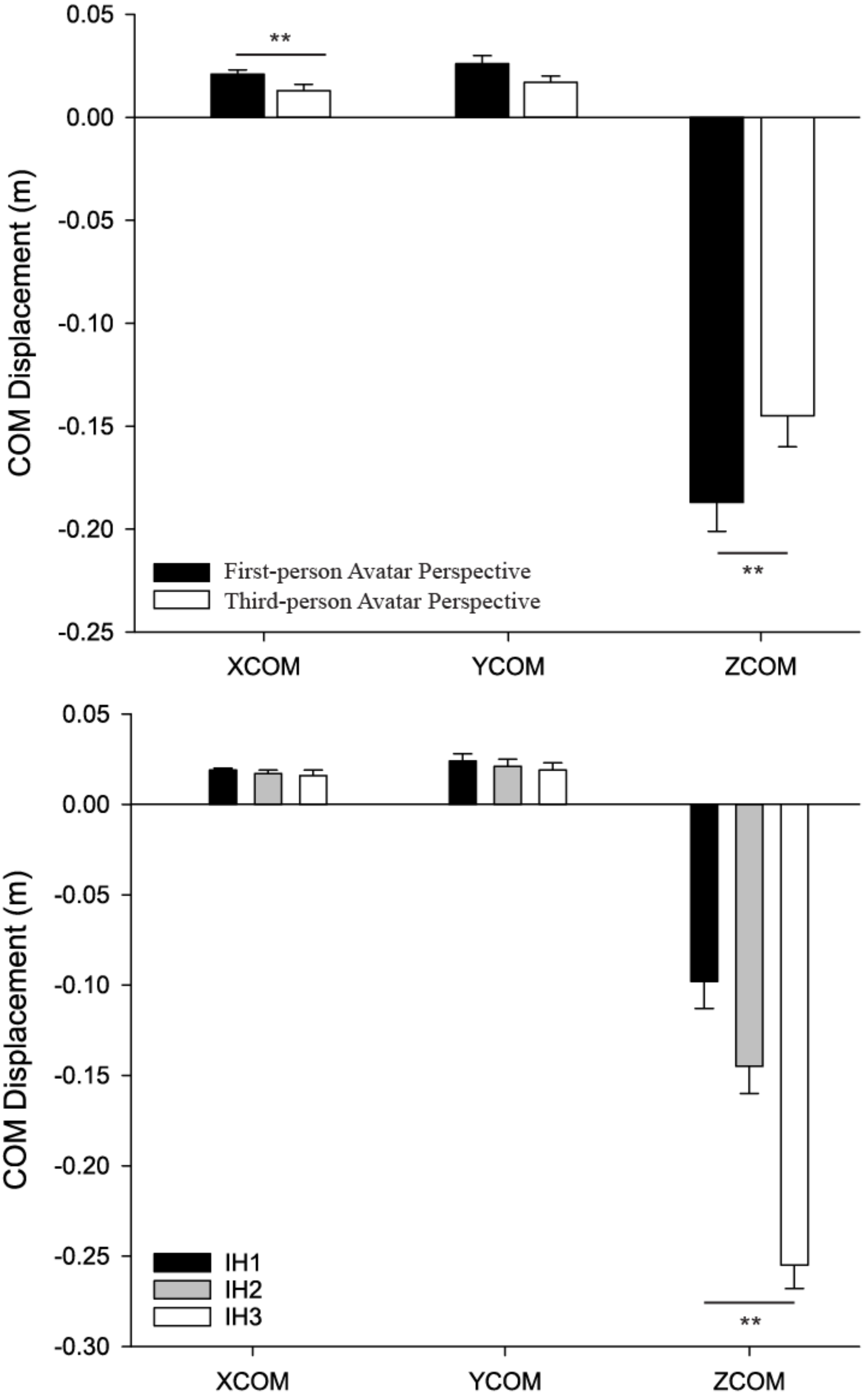
Center of Mass Displacement

Participants had greater forward displacement (along the antero-posterior axis) of their center of mass while playing in a first-person (2.1 cm ± 0.3 cm) versus third-person (1.2 cm ± 0.3 cm) perspective ($F_{1,27}$=16.941, *P*<.001). Impact height also affected antero-posterior displacement of the center of mass ($F_{2,26}$=5.987, *P*=.007), with post-hoc analyses revealing that antero-posterior center of mass displacement was greatest for high (IH1 1.9 cm ± 0.2 cm) versus the middle (IH2 1.7 cm ± 0.3 cm, *P*=.034) and low (IH3 1.5 cm ± 0.3 cm, *P*=.004) impact heights (see Figure 7).

There was an interaction of avatar perspective by impact height on vertical displacement ($F_{2,26}$=17.405, *P*<.001). Follow-up analyses revealed that vertical center of mass displacement was greater when the avatar was presented in first versus the third person for all target heights (IH1 $P<.001$; IH2 $P<.001$; IH3 $P=.002$), but the magnitude of the difference was least for the lowest impact height.

Mediolateral displacement was greater in the first-person perspective (2.8 cm ± 0.5 cm) versus the third-person perspective (1.6 cm ± 0.4 cm), $F_{1,27}=11.432$, $P=.045$. Impact height also affected vertical displacement, $F_{2,26}=3.046$, $P=.045$, reflecting an expected smaller center of mass displacement along the vertical axis for the highest impact height (IH1 2.5 cm ± 0.4 cm) versus the lowest (IH3 2.0 cm ± 0.4 cm, $P=.024$) impact height (see Figure 7).

Figure 7. Effects of avatar perspective (top panel) and : predicted interception heights (bottom panel) on mean COM excursions in the antero-posterior (XCOM), mediolateral (YCOM), and vertical (ZCOM) planes. Error bars represent the standard error of the mean $^{**}P<.001$.



Success

Overall, participants had higher success rates when dodgeball was played in a first-person perspective (95.5%) versus a third-person perspective (91.5%, $F_{1,27}$=5.451, P=.007). The success rate was also influenced by impact height ($F_{2,26}$=6.018, p=.027), with participants being less successful intercepting balls at the low (IH3, 88.4% ± 2.5%) versus middle (IH2, 96.8% ± 1.0%, P=.005) and high (IH1, 95.3% ± 1.1%, P=.010) impact heights.

NASA TLX

Avatar perspective did not affect NASA TLX score or any of its subscales (ie, mental demand, temporal demand, performance, effort, and frustration).

Discussion

Principal Results

The primary goal of this study was to determine the influence of avatar perspective on movement behavior. We achieved this goal by examining changes in joint excursions and center of mass displacement during virtual dodgeball gameplay in a first- and third-person perspective. Consistent with our previous study [8], a first-person avatar perspective resulted in greater joint excursions, center of mass displacement, and hand displacement at ball interception. This finding supports the notion that avatar perspective, rather than simply the mode of displaying the virtual environment and avatar (ie, HMD vs 3DTV), has the potential to influence motor behavior.

The study findings regarding the effect of avatar perspective on motor behavior are consistent with our prior work [8,9] as well an earlier study by Ustinova and colleagues [11], which found participants reached further and less accurately (by measuring the index of curvature of the finger endpoint trajectory throughout the movement) with increasing viewing angles (ie, mid-range) across third-person perspectives. Interestingly, despite these potential decrements in movement efficiency, participants reported a preference for these mid-range viewing angles compared to those with 0° curvature. This view from right behind the avatar may have obscured visual information of the avatar itself, affecting how the body is perceived in space [19] and perception of distance to reach [20]. In the current study, while we did not examine participant preference towards an avatar perspective, we did not find any significant changes in the task loads between first- or third-person perspectives (based on NASA TLX scores); however, participants were more successful and reached further in a first-person compared to third-person perspective.

The differences in joint excursions between first- and third-person perspectives could be explained by changes in how one's avatar and environment are perceived, consistent with theories surrounding movement behavior and embodiment in first- and third-person avatar perspective during observations of avatars in virtual environments [21]. Pavone et al [21] noted that participants report a greater sense of embodiment when an avatar is observed in first- versus third-person perspective, and that embodiment declines when an observed avatar makes grasping errors while the participant is static. Further, errors in the third-person avatar perspective had lower fronto-cortical event-related potentials (triggered by performing an error) as well as medial-frontal theta power (related to action monitoring). Hence, the third-person perspective evoked less of a change in performance-related brain activity as compared to a first-person perspective, indicating errors made in the third-person avatar perspective are less perceived as being one's own errors. However, the difference in success rates between first- and third-person avatar perspectives was significant, suggesting that in both perspectives, success rates were high (95.5% and 91.5%).Thus, avatar errors perceived from a first-person perspective are experienced with a higher sense of embodiment than those in third-person perspective [21]. Therefore, it is likely that third-person avatar perspectives negatively affect the sense of embodiment without many errors, which could be due to the less action monitoring. In the present study, although there was a significant difference in success rates between first- and third-person avatar perspectives (ie, 95.5% and 91.5%, respectively), success rates were nonetheless high in both perspectives. This suggests that a third-person avatar perspective may negatively affect embodiment without producing many errors, which could be due to the less action monitoring (medial-frontal theta power).

Additionally, avatars are perceived more like themselves in the first- than in the third-person perspective [22,23], which could also affect how reaching movements are planned. Altogether, the higher embodiment in the first-person perspective leads to higher success rates but also results in a greater center of mass displacement, and hand and joint excursions. Thus, it could be argued that the movements were more efficient when one's avatar is presented in the third-person perspective. While addressing movement efficiency is beyond the scope of these data, the findings indicated that motor behavior could be altered by manipulating the avatar perspective, which aligns well with our previous work.

The intended impact heights of the launched virtual dodgeballs were calculated based on the participant's trunk length, arm length, and hip height to mimic our standardized full-body reaching paradigm [13,15,16]. While lumbar flexion angles increased significantly from the highest to the lowest targets, the magnitude of lumbar flexion did not match our prior expectations based on the algorithm used to calculate individualized impact heights. Specifically, participants increased joint excursions in a compensatory fashion across several joints, particularly within the lower limb (ie, ankle, knee, hip), leveraging the kinematic redundancy inherent in these whole-body movement tasks. Similar observations have been reported previously using a standardized reaching task as a baseline for target locations [8,14,15]. Only the shoulder joint excursions did not increase across the impact heights; however, a first-person perspective did increase shoulder joint excursion as compared to a third-person perspective. Joint excursions for target reaching tasks have also been shown to be influenced by gender [14], pain-related fear [24], virtual display type [8], and comparison between VR and real-world movements [9]. In the present study, all joints but the elbow showed increased excursions in the first-person perspective as compared to the third-person perspective. The absence of an effect for the elbow joint may be due to the relatively small excursions of this joint required to successfully perform the reaching task.

Finally, the results of this study support virtual dodgeball as an effective strategy to promote lumbar flexion, thereby adding to the clinical utility of virtual dodgeball in promoting joint excursions when the avatar is presented in the first-person perspective. Additionally, similar games can be manipulated for different motions and therefore training different movement behaviors, such as targeting the center of mass motions in the antero-posterior and mediolateral planes to do balance training and kicking motions for control over leg muscle. This approach could provide benefits to populations that require increased motion to alleviate pain (eg, improving movement outcomes for those with chronic low back pain and kinesiophobia). Avatar perspective can thus drive motor behavior in VR gameplay.

Limitations

A limitation of the present study is the absence of a direct comparison with real-world dodgeball. Of course, such a comparison would be a major challenge as real-world dodgeball would be very hard to standardize relative to VR, making it difficult to compare joint, center of mass, and limb excursions. However, simple reaching movements show differences in VR versus real-world whole body movements to the same target height [9]. Therefore it would be expected that if interception positions could be standardized, a similar increase in joint excursions would be elicited in VR compared to real-world full body in virtual dodgeball.

Conclusions

The results of this study demonstrate that the avatar perspective can influence motor behavior. Considering that a primary goal of VR-based rehabilitation is often to restore movement following orthopedic or neurologic injury, understanding how the presentation of an avatar or, by extension, camera position will affect motor behavior is crucial in the development of VR assessment and treatment tools.

Acknowledgments

SMV wrote the manuscript; AS, CRF, and JST proofread the work; SV performed data analysis; MA performed the data collection activities; CRF and JST developed the testing paradgim, project conceptualization, and provided lab space, and funding.

Conflicts of Interest

None declared.

References

1. Thomas JS, France CR, Applegate ME, Leitkam ST, Walkowski S. Feasibility and Safety of a Virtual Reality Dodgeball Intervention for Chronic Low Back Pain: A Randomized Clinical Trial. J Pain 2016 Dec;17(12):1302-1317 [FREE Full text] [doi: 10.1016/j.jpain.2016.08.011] [Medline: 27616607]
2. Fung V, Ho A, Shaffer J, Chung E, Gomez M. Use of Nintendo Wii Fit™ in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial. Physiotherapy 2012 Sep;98(3):183-188. [doi: 10.1016/j.physio.2012.04.001] [Medline: 22898573]
3. Huang M, Lee S, Yeh S, Chan R, Rizzo A, Xu W, et al. Intelligent Frozen Shoulder Rehabilitation. IEEE Intell. Syst 2014 May;29(3):22-28. [doi: 10.1109/mis.2014.35]
4. Subramanian S, Knaut LA, Beaudoin C, McFadyen BJ, Feldman AG, Levin MF. Virtual reality environments for post-stroke arm rehabilitation. J Neuroeng Rehabil 2007 Jun 22;4(1):20 [FREE Full text] [doi: 10.1186/1743-0003-4-20] [Medline: 17587441]
5. Fung J, Richards CL, Malouin F, McFadyen BJ, Lamontagne A. A treadmill and motion coupled virtual reality system for gait training post-stroke. Cyberpsychol Behav 2006 Apr;9(2):157-162. [doi: 10.1089/cpb.2006.9.157] [Medline: 16640470]
6. Magdalon EC, Michaelsen SM, Quevedo AA, Levin MF. Comparison of grasping movements made by healthy subjects in a 3-dimensional immersive virtual versus physical environment. Acta Psychol (Amst) 2011 Sep;138(1):126-134. [doi: 10.1016/j.actpsy.2011.05.015] [Medline: 21684505]
7. Lin H, Li Y, Hu W, Huang C, Du Y. A Scoping Review of The Efficacy of Virtual Reality and Exergaming on Patients of Musculoskeletal System Disorder. J Clin Med 2019 Jun 04;8(6):791 [FREE Full text] [doi: 10.3390/jcm8060791] [Medline: 31167435]
8. Thomas JS, France CR, Applegate ME, Leitkam ST, Pidcoe PE, Walkowski S. Effects of Visual Display on Joint Excursions Used to Play Virtual Dodgeball. JMIR Serious Games 2016 Sep 15;4(2):e16 [FREE Full text] [doi: 10.2196/games.6476] [Medline: 27634561]
9. Thomas JS, France CR, Leitkam ST, Applegate ME, Pidcoe PE, Walkowski S. Effects of Real-World Versus Virtual Environments on Joint Excursions in Full-Body Reaching Tasks. IEEE J. Transl. Eng. Health Med 2016;4:1-8. [doi: 10.1109/jtehm.2016.2623787]
10. Hoffman HG, Doctor JN, Patterson DR, Carrougher GJ, Furness TA. Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. Pain 2000;85(1):305-309. [doi: 10.1016/s0304-3959(99)00275-4]

11. Ustinova K, Perkins J, Szostakowski L, Tamkei L, Leonard W. Effect of viewing angle on arm reaching while standing in a virtual environment: potential for virtual rehabilitation. Acta Psychol (Amst) 2010 Feb;133(2):180-190. [doi: 10.1016/j.actpsy.2009.11.006] [Medline: 20021998]
12. McGill S, Cholewicki J, Peach J. Methodological considerations for using inductive sensors (3space isotrak) to monitor 3-D orthopaedic joint motion. Clinical Biomechanics 1997 Apr;12(3):190-194. [doi: 10.1016/s0268-0033(97)00063-6]
13. Thomas JS, Corcos DM, Hasan Z. Effect of movement speed on limb segment motions for reaching from a standing position. Exp Brain Res 2003 Feb;148(3):377-387. [doi: 10.1007/s00221-002-1287-2] [Medline: 12541148]
14. Thomas JS, Corcos DM, Hasan Z. The influence of gender on spine, hip, knee, and ankle motions during a reaching task. J Mot Behav 1998 Jun;30(2):98-103. [doi: 10.1080/00222899809601327] [Medline: 20037025]
15. Thomas JS, Corcos DM, Hasan Z. Kinematic and kinetic constraints on arm, trunk, and leg segments in target-reaching movements. J Neurophysiol 2005 Jan;93(1):352-364 [FREE Full text] [doi: 10.1152/jn.00582.2004] [Medline: 15342717]
16. Thomas JS, Gibson GE. Coordination and timing of spine and hip joints during full body reaching tasks. Hum Mov Sci 2007 Feb;26(1):124-140 [FREE Full text] [doi: 10.1016/j.humov.2006.08.002] [Medline: 17055098]
17. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016 Nov 05 Presented at: Human Factors and Ergonomics Society Annual Meeting; october 2006; San Francisco p. 904-908 URL: https://journals.sagepub.com/doi/pdf/10.1177/154193120605000909 [doi: 10.1177/154193120605000909]
18. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C (2nd ed.): The art of scientific computing. 40 W. 20 St. New York, NY United States: Cambridge University Press; 1992.
19. Winter DA. Biomechanics and Motor Control of Human Movement. Hoboken, New Jersey: Wiley and sons; 1990.
20. Gianaros PJ, Muth ER, Mordkoff JT, Levine ME, Stern RM. A questionnaire for the assessment of the multiple dimensions of motion sickness. Aviat Space Environ Med 2001 Feb;72(2):115-119 [FREE Full text] [Medline: 11211039]
21. Pavone EF, Tieri G, Rizza G, Tidoni E, Grisoni L, Aglioti SM. Embodying Others in Immersive Virtual Reality: Electro-Cortical Signatures of Monitoring the Errors in the Actions of an Avatar Seen from a First-Person Perspective. J Neurosci 2016 Jan 13;36(2):268-279 [FREE Full text] [doi: 10.1523/JNEUROSCI.0494-15.2016] [Medline: 26758821]
22. Voorhees GA. In: Voorhees GA, Call J, Withlock K, editors. Guns, Grenades, and Grunts First-Person Shooter Games. London UK: Bloomsbury; 2012.
23. Taylor LN. Video games: Perspective, point of view, and immersion. https://www.researchgate.net/publication/35486642_Video_games_electronic_resource_perspective_point-of-view_and_immersion. 2002. URL: https://www.researchgate.net/publication/35486642_Video_games_electronic_resource_perspective_point-of-view_and_immersion [accessed 2020-08-06]
24. Thomas JS, France CR, Sha D, Wiele NV. The Influence of Pain-Related Fear on Peak Muscle Activity and Force Generation During Maximal Isometric Trunk Exertions. Spine 2008;33(11):E342-E348. [doi: 10.1097/brs.0b013e3181719264]

Abbreviations

HMD: head-mounted display
TLX: Task load index
TV: television
VR: Virtual reality
3D: Three-dimensional

Edited by G Eysenbach; submitted 25.03.20; peer-reviewed by L Sheehy, J Salisbury; comments to author 12.06.20; revised version received 24.06.20; accepted 07.07.20; published 19.08.20.

Please cite as:
van der Veen SM, Stamenkovic A, Applegate ME, Leitkam ST, France CR, Thomas JS
Effects of Avatar Perspective on Joint Excursions Used to Play Virtual Dodgeball: Within-Subject Comparative Study
JMIR Serious Games 2020;8(3):e18888
URL: http://games.jmir.org/2020/3/e18888/
doi:10.2196/18888
PMID:32812885

Review

Gamification in Rehabilitation of Patients With Musculoskeletal Diseases of the Shoulder: Scoping Review

Bianca Steiner[1], MSc; Lena Elgert[2], MSc; Birgit Saalfeld[2], MSc; Klaus-Hendrik Wolf[2], PhD

[1]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany

[2]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany

Corresponding Author:
Bianca Steiner, MSc
Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School
Muehlenpfordtstr 23
Braunschweig, 38106
Germany
Phone: 49 531 391 9518
Email: bianca.steiner@plri.de

Abstract

Background: Gamification has become increasingly important both in research and in practice. Particularly in long-term care processes, such as rehabilitation, playful concepts are gaining in importance to increase motivation and adherence. In addition to neurological diseases, this also affects the treatment of patients with musculoskeletal diseases such as shoulder disorders. Although it would be important to assist patients during more than one rehabilitation phase, it is hypothesized that existing systems only support a single phase. It is also unclear which game design elements are currently used in this context and how they are combined to achieve optimal positive effects on motivation.

Objective: This scoping review aims to identify and analyze information and communication technologies that use game design elements to support the rehabilitation processes of patients with musculoskeletal diseases of the shoulder. The state of the art with regard to fields of application, game design elements, and motivation concepts will be determined.

Methods: We conducted a scoping review to identify relevant application systems. The search was performed in 3 literature databases: PubMed, IEEE Xplore, and Scopus. Following the PICO (population, intervention, comparison, outcome) framework, keywords and Medical Subject Headings for shoulder, rehabilitation, and gamification were derived to define a suitable search term. Two independent reviewers, a physical therapist and a medical informatician, completed the search as specified by the search strategy. There was no restriction on year of publication. Data synthesis was done by deductive-inductive coding based on qualitative content analysis.

Results: A total of 1994 articles were screened; 31 articles in English, published between 2006 and 2019, were included. Within, 27 application systems that support patients with musculoskeletal diseases of the shoulder in exercising, usually at home but also in inpatient or outpatient rehabilitation clinics, were described. Only 2 application systems carried out monitoring of adherence. Almost all were based on in-house developed software. The most frequently used game components were points, tasks, and avatars. More complex game components, such as collections and teams, were rarely used. When selecting game components, patient-specific characteristics, such as age and gender, were only considered in 2 application systems. Most were described as motivating, though an evaluation of motivational effects was usually not conducted.

Conclusions: There are only a few application systems supporting patients with musculoskeletal diseases of the shoulder in rehabilitation by using game design elements. Almost all application systems are exergames for supporting self-exercising. Application systems for multiple rehabilitation phases seem to be nonexistent. It is also evident that only a few complex game design elements are used. Patient-specific characteristic are generally neglected when selecting and implementing game components. Consequently, a holistic approach to enhance adherence to rehabilitation is required supporting patients during the entire rehabilitation process by providing motivational game design elements based on patient-specific characteristics.

(JMIR Serious Games 2020;8(3):e19914) doi:10.2196/19914

KEYWORDS

shoulder; upper extremity; musculoskeletal diseases; rehabilitation; gamification; serious games; exergames; scoping review

Introduction

Background

Musculoskeletal diseases are one of the leading causes of chronic joint pain and physical disability worldwide [1]. In Germany, the prevalence of chronic pain is approximately 17%, depending on the affected joint [2]. Thereby, chronic shoulder pain belongs to the most common forms of musculoskeletal diseases, which lead to high socioeconomic costs [2]. The most common causes for chronic shoulder pain are shoulder lesions, for example frozen shoulder (6%), osteoarthritis (5%-10%) and rotator cuff tears (10%) [3]. In addition to intense shoulder pain, affected persons also suffer from disabilities in shoulder mobility and functionality [4]. Apart from pain-reducing drugs and surgical treatment, rehabilitative procedures are part of the standard therapy [5]. Rehabilitation is a multifaceted long-term process ranging from inpatient or outpatient orthopedic rehabilitation up to subsequent rehabilitation services [6]. Thereby, the conservative treatment includes a multitude of interventions, such as physical therapy, exercise therapy, psychological treatment, naturopathy, alternative medicine, as well as activities such as swimming or yoga [7,8]. To maintain or improve the success of therapy in patients with musculoskeletal diseases of the shoulder, an ongoing execution of acquired changes in behavior and lifestyle, as well as a long-term provision of subsequent rehabilitation services is required [9]. Accordingly, patients' motivation and adherence are crucial factors for effective rehabilitation [10]. However, only 50% of all patients suffering from chronic diseases achieve good adherence [11]. Although therapeutic adherence is defined as "the extent to which a person's behavior – taking medication, following a diet, or executing lifestyle changes, corresponds with agreed recommendations from health care provider [11]," it is not only about following recommendations of physicians or therapists but rather about following general measures to achieve individual therapeutic goals [12]. A multitude of factors influence a patient's adherence, both in a positive and a negative way [13].

Today's situation is characterized by a high number of patients with musculoskeletal diseases of the shoulder, low adherence to long-term therapies, and a resulting enormous financial burden on health care systems. This motivates finding modern solutions for supporting patients during rehabilitation and thus reducing costs. Among other things, health-enabling technologies are being developed, which are defined as "sensor-based information and communication technologies, aiming at contributing to a person's health and health care as well as to his or her quality of life [14]." An increasing number of such health-enabling technologies are supplemented by playful concepts. Particularly in long-term processes, gamification is gaining in importance to increase intrinsic motivation and adherence [15], whereby various game design elements, meaning game components, game dynamics, and game mechanics, are used to generate playfulness [16]. Although it would be important to assist patients during more than one rehabilitation phase, it is hypothesized that existing application systems only support a single phase. It is also unclear which game design elements are currently used in this context and how they are combined to achieve optimal positive effects on motivation and adherence.

Objectives

Within this scoping review, information and communication technologies utilizing game design elements will be identified to obtain an overview of the fields of application, the game design elements, and motivation concepts. Accordingly, 3 categories should be analyzed: *rehabilitation process*, *gamification*, and *motivation*. As part of these, this review attempts to answer the following questions:

1. Rehabilitation Process

- Scope: What is the scope of the identified application systems?
- Phase: In which phase of rehabilitation processes are the identified application systems used?

2. Gamification

- General implementation: How is the concept of gamification generally implemented in the application systems?
- Game design elements: Which game design elements, that is game components, mechanics, and dynamics, are used?

3. Motivation

- Motivation concepts: Which motivation concepts are addressed?
- Evaluation and outcome: Which aspects of the application systems are evaluated, in particular, motivation and adherence?

Methods

Overview

The overall aim of this review was to identify and analyze existing technology-based approaches using game design elements to support rehabilitation processes of patients with musculoskeletal diseases of the shoulder. For this purpose, this scoping review was conducted in accordance with the extended PRISMA checklist for scoping reviews (PRISMA-ScR) [17]. Given the exploratory research question, a scoping review seemed to be more appropriate and thus more effective than a systematic review to get an overview of the field of application and derive appropriate hypotheses.

Eligibility Criteria

In general, peer-reviewed articles, reviews, and conference articles published between 1997 and mid 2019 were included. For literature not written in English or German, a translator was involved. If, despite all efforts, it was not possible to translate a relevant publication, it was excluded retrospectively. Reviews were only used additionally as part of a citation pearl growing.

The eligibility criteria were determined on basis of the PICO (population, intervention, comparison, outcome) framework [18]; however, *outcome* and *comparison* were not considered. In terms of *population*, all publications focusing on patients with musculoskeletal diseases of the shoulder joint were included. This covered patients with following conditions: fractures, dislocations, inflammatory diseases, and degenerative diseases [19]. Due to treatment features, musculoskeletal diseases such as tumors and tumor-like lesions, congenital diseases of the skeletal system, as well as diseases of the child's skeletal system were not considered. In terms of *intervention*, all publications describing an application system used to support patients within the rehabilitation process were included. It should be noted that only application systems in which game design elements were used, that is, game components, game mechanics, and game dynamics, were included. Hardware-based systems such as robots, exoskeletons, and orthoses were not considered.

Information Sources

For a comprehensive overview, we searched 3 databases: the medical literature database Medline via PubMed, the technical research database IEEE Xplore, and the multidisciplinary database Scopus.

Search Strategy

A proper search term was defined iteratively following PICO. According to the objectives, the key elements *population* (population, patient, problem) and *intervention* were crucial for this review. Patients with musculoskeletal diseases of the shoulder are described by various terms (keywords and Medical Subject Headings) for the shoulder or upper extremities (eg, shoulder, arm, upper limb, or glenohumeral). There was no specific search for musculoskeletal diseases of the shoulder, as it was assumed that these are rarely mentioned explicitly. Intervention describes a variety of different interventions, all of which are subsumed under the term *rehabilitation*. Rehabilitation ranges from diagnosis and acute treatment to therapy planning, medical, or vocational rehabilitation to subsequent rehabilitation services. Therefore, apart from general terms such as *rehabilitation*, *computer*, or *technology-assisted therapy*, more specific therapeutic terms were searched (eg, *training*, *physical therapy*, or *occupational therapy*). Additional terms, such as *therapy planning* or *self-management*, were added to allow a more comprehensive search. Technical aspects were considered by a fourth search block including both specific terms such as *game*, *gamification*, and *exergame*, as well as unspecific terms such as *virtual* and *augmented reality*. For more details on the search queries see Multimedia Appendix 1.

Synthesis of Results

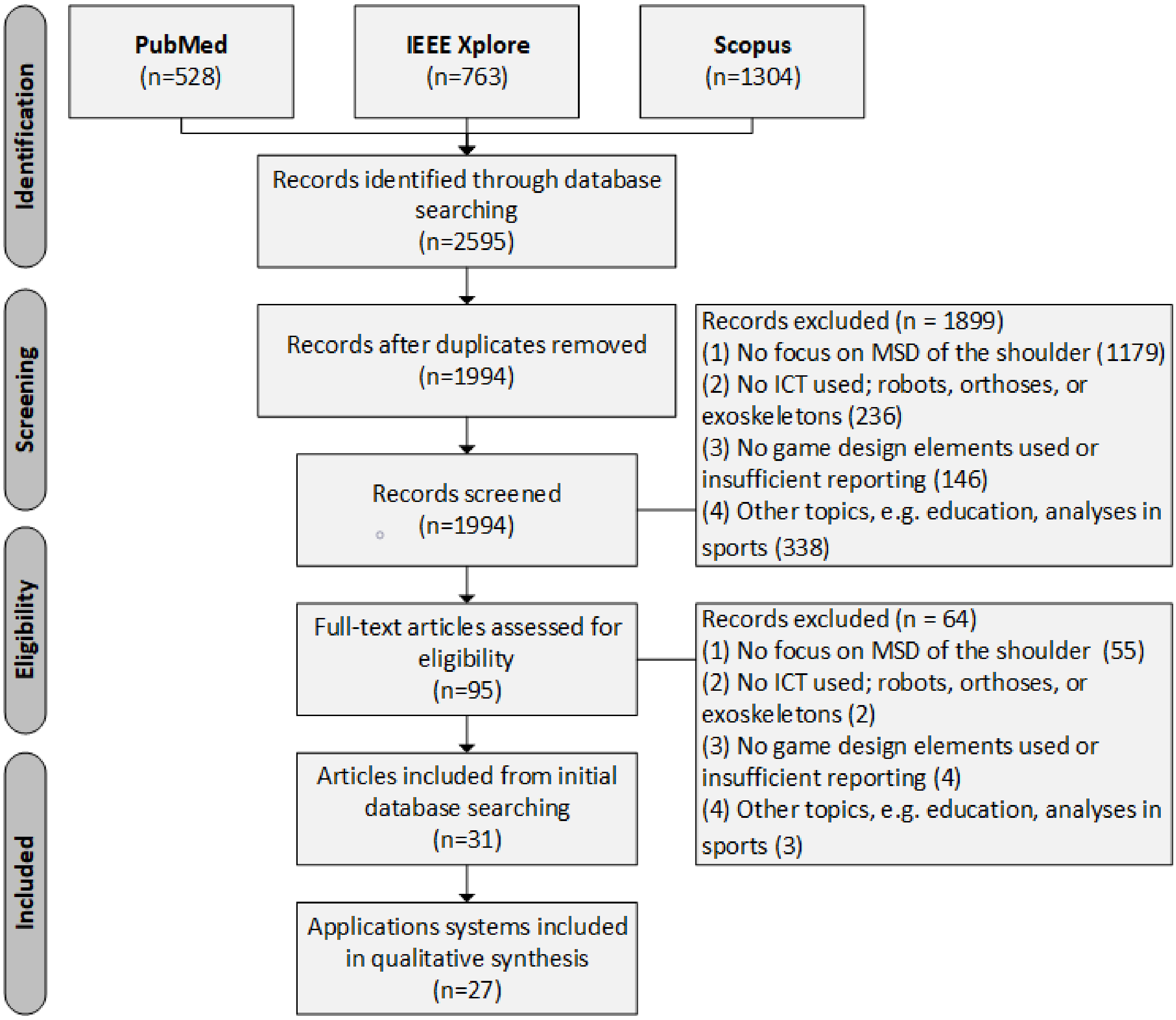
Following qualitative content analysis [20], full texts classified as relevant were summarized, analyzed, and evaluated. To answer the predefined questions adequately, the basic process was adapted by using the toolbox for qualitative content analysis based on Schreier [21]. Here, an essential part was the deductive-inductive approach for creating a category system. The starting point was a code tree consisting of 16 deductively built subcategories and 6 main categories: (1) target group, (2) field of application, (3) rehabilitation phase, (4) motivation aspects, (5) game characteristics, and (6) evaluation. The applied methodology of mixed deductive-inductive categorization allowed a quantitative evaluation in addition to a purely qualitative analysis of publications [20]. In order to simplify and partially automate the analysis, qualitative data analysis software MAXQDA (version 2018.2; VERBI GmbH) was used.

Selection, Categorization, and Data Extraction

The search request was made to PubMed on June 17, 2019 and to IEEE Xplore and Scopus on June 18, 2019. Overall, the search returned 2595 matches. After excluding duplicates, title screening and abstract screening of the remaining 1994 publications were conducted by 2 independent reviewers (LE, BSt) with different expertise—physical therapy and medical informatics. Checking of predefined exclusion criteria was done sequentially. Whenever disagreements arose, a third reviewer (BSa) was consulted for decision making. Thus, a total of 95 publications were included in the full-text screening. Full-text screening was carried out in the same way as the title and abstract screening. Finally, 31 publications were included in the full-text analysis (Figure 1).

Qualitative content analysis was performed by a single reviewer (BSt) using the predefined search tree. During the analysis, 17 inductive subcategories were added (see Multimedia Appendix 2 for more details).

Figure 1. PRISMA Flow diagram for the identification of application systems. ICT: information communication technology; MSD: musculoskeletal diseases.



Results

Overview

A total of 31 articles published between 2006 and 2019 were selected for full-text analysis. These articles introduced 27 different application systems to support the rehabilitation of patients with musculoskeletal diseases of the shoulder by using game design elements. A list of identified articles grouped by the underlying population can be found in Multimedia Appendix 3, and a complete overview of all analyzed aspects can be found in Multimedia Appendix 4.

Population

Of the 27 application systems, 10 had been developed explicitly for a musculoskeletal disease of the shoulder—6 for frozen shoulder, 4 for shoulder impingement syndrome; 2 focused on different musculoskeletal diseases of the shoulder similar in physiotherapeutic treatment (eg, rotator cuff tear and humerus fracture); 11 others outlined the target group only vaguely—there were 5 application systems for shoulder injuries, 2 for musculoskeletal shoulder pain, and 4 for musculoskeletal diseases of upper extremities, in general; and the remaining 4 were intended to be used for treatment of musculoskeletal diseases of the shoulder as well as of one of the following diseases: paraplegias with spinal cord injuries [22], cerebrovascular diseases [23], arm amputations [24], or elbow and radius fractures [25]. Here, it seems questionable whether a combined approach for musculoskeletal and neurological diseases is appropriate for the individual therapeutic needs of a patient. While patients with such neurological diseases may also suffer from pain and limited range of motion of upper extremities [26-28], the causes and symptoms, the patients' self-management skills, and the therapeutic goals may differ. Especially for severely affected patients, there are nuances in the therapy plans, exercise programs, and ultimately, also in the physiotherapeutic exercises to be performed.

Rehabilitation Process

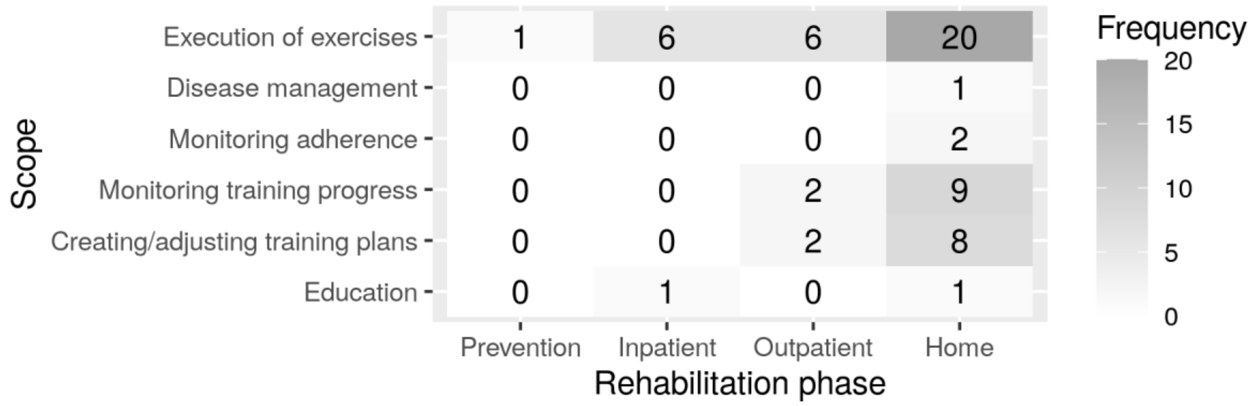
Scope

All 27 application systems assist in the execution of exercises, mostly physiotherapeutic self-exercises. Of them, 5 were explicitly described as telerehabilitation systems [25,29-32]. These application systems provide additional functionalities for creating and adjusting individual exercises and training plans or for monitoring training progress by a therapist or physician; however, only Rahman et al [33] and Luchessi et al [30] describe possibilities for monitoring patients' adherence. The application system GEAR, for example, has a dashboard for physicians to monitor adherence by visualizing the exercise performance in different diagrams [29,33]. Beyond this, the app SHOULPHY claimed to offer possibilities for disease management [30]. Given the existing definitions of disease management [34,35], this does not fully apply. Although patient empowerment is increased and physicians are supported in the development and maintenance of a treatment plan [34], there is no holistic, cross-process view on care over the entire life cycle of the disease [35].

Phase

Most of the application systems were developed for the outpatient sector as part of conservative therapy or subsequent rehabilitation services (Figure 2). Only Mangal et al [36] addressed the additional employment of a Kinect-based rehabilitation system for treatment of frozen shoulder in inpatient rehabilitation. Vogt et al [37] and Wiederhold et al [24] add 2 training systems explicitly developed for inpatient rehabilitation. PhysioSonic, for example, is a sonification system for shoulder patients installed at an orthopedic hospital [37]. Besides the multitude of training systems for later rehabilitation phases, Huang et al [38] are the only ones presenting an app for prevention. Using a bicycle ergometer, Google Earth, and a Kinect camera, exercises based on pilates can be performed.

Figure 2. Heatmap of frequencies of scopes in individual rehabilitation phases; outpatient and inpatient refer to a rehabilitation facility (multiple mentions are possible in both axes).



Gamification

General Implementation

Of the 27 application systems, 24 were based on self-developed software products. Only 3 made use of commercial games, namely Wii Sports [39,40]; EyeToy: Play, Dance Dance Revolution; and Taiko Drum Master [24]. While commercial games do not require any investments for development, in contrast to in-house developed games, they are not specially adapted to the target group. In terms of the selection and implementation of exercises and game design elements, it would be necessary to include patient-specific characteristics. However, only 2 consider either of these characteristics, that is, age and gender [41,42].

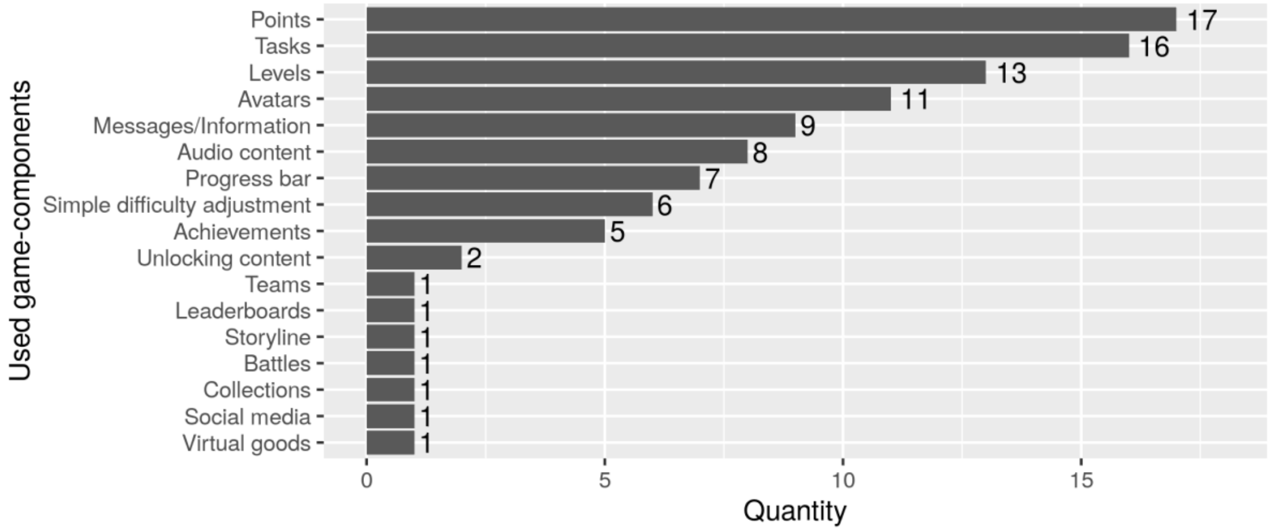
Whereas 11 publications focus on the implementation of a single, sometimes incredibly broad game [22,23,29,37,38,43-48], just as many publications provide several games [24,28,32,36,39-42,49-51]. By implementing different games, a higher variability is gained. Thus, motivation and acceptance of users can be additionally increased by avoiding boredom. Furthermore, it is possible to match therapeutic interventions much better to the patients' state of health. A user's personal preferences can be also addressed, even if only indirectly. For example, in the application system InMotion users can choose to touch butterflies in a beautiful garden, hit and shoot paintballs towards a canvas, control a rowboat through several checkpoints, or mimic the motions of a pendulum [42].

Game Components

The most frequently used game components are those that are easy to implement, such as points, tasks, avatars, and information messages (Figure 3). When playing Cupid's Arrow, for example, players collect points by shooting an arrow at a heart through correct arm movements [41]. Also, in Classic Clock, points can be earned by imitating movements of a pendulum as precisely as possible [42]. However, points can also be used in application systems implementing no specific game. Da Gama et al [52] demonstrate how points can be gained by completing physiotherapeutic exercises correctly, thus boosting the total score of a therapy session.

Most application systems implement training as playful tasks. Apart from being more fun, this is more interesting than doing the same monotonous exercises on a daily basis [44]. In addition to simple tasks such as picking fruits [25,47,50,53], placing dishes [50-51], or fishing [44], there are more complex, and therefore, possibly more exciting and less fatiguing tasks. Examples are destroying submarines by cleverly dropping bombs [41] and controlling objects (eg, aircrafts, storks, rays) along a course with or without obstacles [23,25,32,38]. Probably the most extensive game in terms of implemented tasks and game components is The Sorcerer's Apprentice [45]. Here, a mage (avatar) must be controlled in a virtual world (storyline). Artefacts must be collected (tasks, virtual goods) to unlock and execute different shoulder exercises (unlocking content). Successfully completed exercises are rewarded with items and bonus points (collection, points).

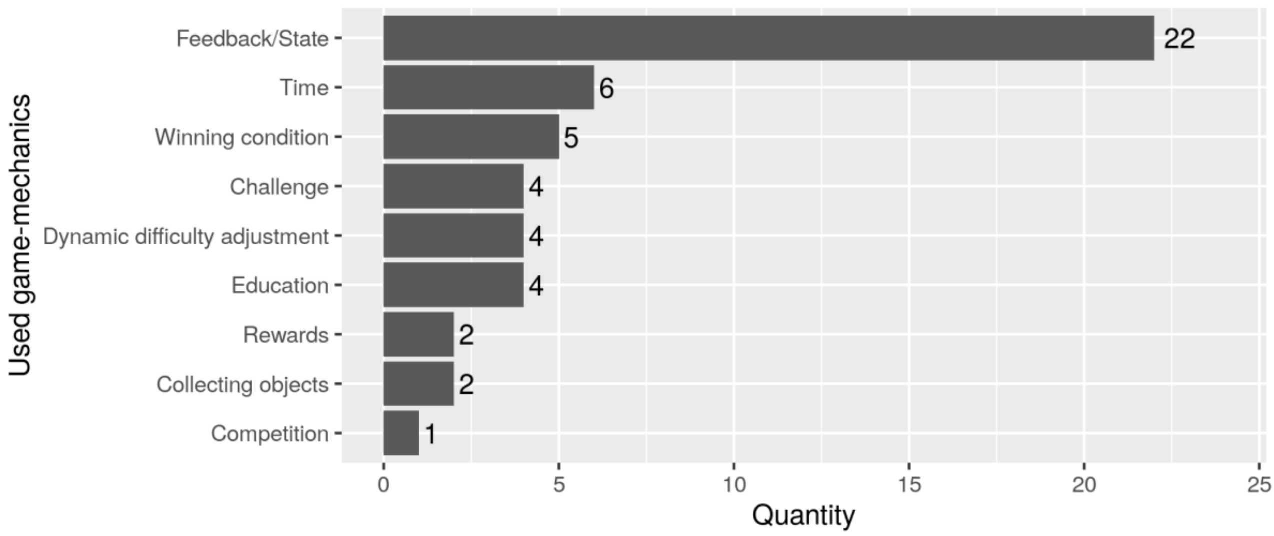
Figure 3. Overview of game components ordered by frequency.



Game Mechanics

Nearly all application systems used game components, such as points, progress bars, leaderboards, and achievements to provide feedback or represent a current state (Figure 4). Since all application systems were implemented to support training, half do not give feedback on the game (game feedback) but instead give feedback on the performance of exercises (therapy feedback). Therapy feedback often quantifies either the quality of exercise performance [37,42,45,46,52] or the progress/current status [44,50] to motivate patients or to ensure correct practice. Almost all application systems provided instructions, or even complete tutorials, for exercising correctly. Nevertheless, there were few providing additional information or education about rehabilitation. Only the exergame by Nava et al [48] made information on the relevance of continuous training available. More complex game mechanics were rarely implemented. Cooperation, round-based games, transaction, and coincidence were not found at all.

Figure 4. Overview of game mechanics ordered by frequency.



Game Dynamics

According to García-Martínez et al [54], all implemented tasks are based on the same 6 game dynamics (activities) for therapeutic games: (1) follow the path, (2) touch the target, (3) point and shoot the target, (4) free movement, (5) move the target, and (6) catch the target. Most commonly used are follow the path (12 of the 27) and touch the target (11 of the 27), as this is closest to normal, conscious movements of exercise therapy for the shoulder. Activities such as free movement, move the target, and catch the target, on the other hand, are used less frequently, as they can lead more quickly to incorrect movements and thus injuries or overstraining [55]. Quite often, a combination of different movements, and thus game dynamics, takes place to motivate users and to respond to different therapy needs.

Motivation

Motivation Concepts

Encouraging motivation, as well as enhancing engagement, is one of several goals of gamification [56,57]. Certain publications describe their application systems as motivating (6 of the 27), without mentioning specific factors for this effect [23,25,32,36,39,46]. Furthermore, motivational effects are often based on general, superficial aspects, such as interactivity (12 of the 27), fun, enjoyment, and entertainment (13 of the 27). In some cases, there are also more concrete aspects mentioned, such as engagement (11 of the 27) and avoiding boredom (10 of the 27). Just 4 publications addressed the need for a better understanding of the exercises to be performed, in terms of the therapeutic goals [37,44] and the associated faster learning of exercises [42,53]. Considering established motivation models, for instance the self-determination theory of motivation [58], it becomes clear how these aspects represent an elementary factor for intrinsic motivation.

Evaluation and Outcome

Except for 3 publications presenting purely application systems or approaches [23,47,59], all application systems were evaluated on at least one aspect. Frequently, this was the evaluation of technical feasibility (10 of the 27) and usability (9 of the 27) by means of pilot studies or simple informal pretests. Resulting from the use of gamification, motivation is another frequently analyzed criterion (13 of the 27). This is described either as intrinsic motivation, motivation to perform exercises, or motivation to regularly perform therapy sessions. Of 14 application systems, 5 were rated as motivating in pretests or pilot studies [22,29,36,48,52]. However, it is unclear to what extent these results are transferable to the population owing to small sample sizes [29,52], healthy people as users [29,36,48,52], or informal testing [22]. The remaining application systems simply suppose an increase in motivation (9 of the 27)—motivating effects were usually deduced from single subjective questions [36-38,42] or directly assumed without any evidence as a result of the game design elements used [24,30,40,41,45]. The same applied to adherence analyzes; 3 publications presume an increase was caused by the game design elements that had been implemented, although there was no evidence given [29,30,40].

Discussion

Principal Findings

A previous review [60] has already shown the high number of application systems available for assisting patients with neurological disorders during rehabilitation, varying from Parkinson disease, stroke, and cerebral palsy. For rehabilitation of musculoskeletal diseases, in contrast, only a few publications can be found. This scoping review identified a total of 27 application systems to support patients with musculoskeletal diseases of the shoulder during their rehabilitation by using game design elements. It is remarkable that only about one-third of these application systems were designed for a specific musculoskeletal disease of the shoulder. Musculoskeletal diseases of the shoulder, such as shoulder dislocation or humeral head fracture, are not given special consideration.

The few application systems that can be found are mostly exergames supporting physiotherapeutic self-exercises, usually at home (20 of the 27). Application systems to assist prevention, diagnostics, acute treatment, or inpatient rehabilitation seems to be almost nonexistent. The same applies to application systems supporting multiple rehabilitation phases. Here, not even one could be discovered.

There is also little variation in the scope of application systems. All application systems support the training of patients, but only a few can be used to create training plans (8 of the 27), or even, to adjust them during the rehabilitation process (6 of the 27). Also, functionalities for monitoring therapy adherence and progress are quite rare (9 of the 27). Just 2 offer possibilities for monitoring therapy adherence, although adherence is a crucial factor for sustainable rehabilitation success [10], whereby an adequate management of the underlying musculoskeletal diseases is necessary to achieve adherence throughout the entire lifecycle of the disease. Luchessi et al [30] are the only ones who actually mention the term *disease management*.

Almost all application systems are based on in-house developed software (24 of the 27). They offer flexibility in terms of the therapeutic needs of patients with shoulder diseases as well as of the game design elements used to enhance motivation and adherence. Selection and implementation of adequate game design elements depend substantially on the target group and its individual characteristics, such as age, gender, personal preferences, level of knowledge, and intrinsic motivation [15,42]. However, the identified application systems generally describe the user group exclusively by indication for treatment. Only 2 approaches address gender and age differences. Other patient-specific characteristics have not yet been considered in selecting nor in implementing game design elements.

Analyzing the application systems in terms of the game design elements used, it is obvious that more complex game components are used only rarely or not at all. Usually, easy to implement game components are selected and combined. Since the realization of game mechanics is based on the game components implemented, even here, little variation is found. Accordingly, 22 application systems used diverse simple game components to provide feedback. Astonishingly, education was also rarely included, although this is a meaningful aspect for patient empowerment. Overall, it seems that only little attention has been paid to the factual effect of individual game components and game mechanics when selecting and implementing them. No publications deal with respective motivation models and theories to demonstrate or prove the effects of single game design elements on user behavior.

Almost all application systems had been evaluated. Mostly, this concerned the technical feasibility, efficacy, or acceptance but rarely the motivation and adherence of patients. Factual evidence of increased motivation or adherence caused by the developed application systems or game design elements as part of clinical trials was lacking.

Limitations

There were only a few application systems for the rehabilitation of patients with musculoskeletal diseases of the shoulder using game design elements. Excluding hardware-based systems limited the search considerably. The same applied to the exclusion of publications due to insufficient descriptions of game design elements during title and abstract screening. Here, numerous publications were excluded (n=150). Given an inadequate description of game design elements in the abstract, it is not possible to determine for certain that such elements had not been implemented. However, this indicated that gamification does not have a central role in these publications. For analyses with regard to the motivational effects of game design elements, this is nevertheless, an essential prerequisite.

Whereas all screenings were performed by 2 independent reviewers, the qualitative content analysis was done by a single reviewer. A second reviewer might have ensured the correct identification and classification of relevant game design elements, rehabilitation phases, and motivation aspects. Particularly in classifying game mechanics and game dynamics, there is some margin for interpretation. Consequently, it was not possible to provide details on game dynamics in the sense of conceptual structures, such as constraints, emotions, or relationships [61]. However, to avoid incorrect classifications, in case of hesitation, one of the other reviewers was always consulted.

Overall, it is striking that only exergames supporting outpatient or inpatient rehabilitation could be identified. There was just one application system for prevention and one for disease management. This restricted result might be caused by the common meaning of rehabilitation as medical and vocational rehabilitation including subsequent rehabilitation services. Even searching for terms such as *self-management*, *prevention*, and *patient pathway* was not sufficient. Therefore, a complementary mobile health app analysis is planned using the Google Play Store to examine whether there are other application systems beyond exergames and subsequent rehabilitation services for musculoskeletal diseases of the shoulder.

Conclusions

The concept of gamification is gaining importance in the context of health care to enhance motivation and support therapy in general. Especially for chronic diseases and thus long-term care processes such as rehabilitation, gamification reveals particular potential. However, for rehabilitation of musculoskeletal diseases of the shoulder, only a few application systems that use game design elements to increase motivation and adherence exist. Beyond that, all identified application systems focus on the support of self-exercises, mainly in the outpatient sector. Application systems for inpatient rehabilitation, other rehabilitation phases, or multiple rehabilitation phases seem to be nonexistent. Only the fewest application systems provide additional functionalities, for example, provision of information or monitoring of adherence to assist patients during rehabilitation and increase their self-management skills. Altogether, the selection, combination, and implementation of game design elements appears somewhat impetuous. Indeed, gamification is generally used to motivate, avoid boredom, and distract from pain and anxiety. But it seems that only a little attention has been paid to the factual effect and thus benefit of individual game components and game mechanics. Furthermore, patient-specific characteristics are mostly neglected when selecting game components. It remains exciting to learn about the effects that future developments will report, including patient-specific game design elements.

Acknowledgments

Open access funding was provided by the German Research Foundation and the Open Access Publication Funds of the *Technische Universität Braunschweig*.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries.
[PDF File (Adobe PDF File), 34 KB - games_v8i3e19914_app1.pdf]

Multimedia Appendix 2
Code tree.
[PDF File (Adobe PDF File), 96 KB - games_v8i3e19914_app2.pdf]

Multimedia Appendix 3
Search results grouped by the underlying population.
[PDF File (Adobe PDF File), 93 KB - games_v8i3e19914_app3.pdf]

Multimedia Appendix 4
Complete overview of search results.

[XLSX File (Microsoft Excel File), 17 KB - games_v8i3e19914_app4.xlsx]

References

1. Woolf AD, Erwin J, March L. The need to address the burden of musculoskeletal conditions. Best Pract Res Clin Rheumatol 2012 Apr;26(2):183-224. [doi: 10.1016/j.berh.2012.03.005] [Medline: 22794094]
2. Wolff R, Clar C, Lerch C, Kleijnen J. [Epidemiology of chronic non-malignant pain in Germany]. Schmerz 2011 Feb;25(1):26-44. [doi: 10.1007/s00482-010-1011-2] [Medline: 21336883]
3. Mitchell C, Adebajo A, Hay E, Carr A. Shoulder pain: diagnosis and management in primary care. BMJ 2005 Nov 12;331(7525):1124-1128 [FREE Full text] [doi: 10.1136/bmj.331.7525.1124] [Medline: 16282408]
4. Fuchs J, Rabenberg M, Scheidt-Nave C. [Prevalence of selected musculoskeletal conditions in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May;56(5-6):678-686. [doi: 10.1007/s00103-013-1687-4] [Medline: 23703486]
5. Beck L, Brüggemann S, Druckenmüll A, Falk J, Fecher K, Flohr S, et al. Reha-Bericht 2018: Die medizinische und berufliche Rehabilitation der Rentenversicherung im Licht der Statistik. Berlin: Deutsche Rentenversicherung Bund; 2018 Jan. URL: www.bdpk.de/fileadmin/user_upload/BDPK/News/2018/PDF/DRV_Rehabericht_2018.pdf [accessed 2020-08-13]
6. Holmes RE, Barfield WR, Woolf SK. Clinical evaluation of nonarthritic shoulder pain: Diagnosis and treatment. Phys Sportsmed 2015 Jul;43(3):262-268. [doi: 10.1080/00913847.2015.1005542] [Medline: 25622930]
7. Greitemann B, Stein V. Praxis der konservativen Orthopädie – Rehabilitation in der Orthopädie. Orthopädie und Unfallchirurgie up2date 2013 Dec 10;8(06):503-522. [doi: 10.1055/s-0033-1346912]
8. Chaitow L. Naturopathic Physical Medicine: Theory and Practice for Manual Therapists and Naturopaths. Edinburgh: Elsevier Health Sciences; 2008.
9. Bundesarbeitsgemeinschaft für Rehabilitation. Praxisleitfaden: Strategien zur Sicherung der Nachhaltigkeit von Leistungen zur medizinischen Rehabilitation. Frankfurt/Main: BAR e.V; Feb 2008.
10. Dunbar-Jacob J, Erlen JA, Schlenk EA, Ryan CM, Sereika SM, Doswell WM. Adherence in chronic disease. Annu Rev Nurs Res 2000;18:48-90. [Medline: 10918932]
11. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva: World Health Organization; 2003.
12. Koch U, Bengel J. Therapieadhärenz. In: Enzyklopädie der Psychologie. Anwendungen der Medizinischen Psychologie. Göttingen: Hogrefe Verlag; 2017:527-529.
13. Mathes T, Jaschinski T, Pieper D. Adherence influencing factors - a systematic review of systematic reviews. Arch Public Health 2014;72(1):37 [FREE Full text] [doi: 10.1186/2049-3258-72-37] [Medline: 25671110]
14. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: Review of the State-of-the-Art and Areas of Application. JMIR Rehabil Assist Technol 2017 Jul 21;4(2):e7 [FREE Full text] [doi: 10.2196/rehab.7511] [Medline: 28733271]
15. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
16. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining Gamification. New York: Association for Computing Machinery; 2011 Presented at: MindTrek '11: 15th International Academic MindTrek Conference: Envisioning Future Media Environments; 2011; Tampere Finland. [doi: 10.1145/2181037.2181040]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Hoogendam A, de Vries Robbé PF, Overbeke AJPM. Comparing patient characteristics, type of intervention, control, and outcome (PICO) queries with unguided searching: a randomized controlled crossover trial. J Med Libr Assoc 2012 Apr;100(2):121-126 [FREE Full text] [doi: 10.3163/1536-5050.100.2.010] [Medline: 22514508]
19. Linaker CH, Walker-Bone K. Shoulder disorders and occupation. Best Pract Res Clin Rheumatol 2015 Jun;29(3):405-423 [FREE Full text] [doi: 10.1016/j.berh.2015.04.001] [Medline: 26612238]
20. Mayring P. Qualitative Inhaltsanalyse. In: Mey G, Mruck K, editors. Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: VS Verlag für Sozialwissenschaften, Springer Fachmedien; 2010:601-613.
21. Schreier M. Varianten qualitativer Inhaltsanalysein Wegweiser im Dickicht der Begrifflichkeiten. Forum Qualitative Sozialforschung 2014 Jan;15(1).
22. Shroeder B, Kunanec M, Kroese B, Pollard O, Sadoon O, Kapralos B. Rapid Recovery: A Kayaking-Based Exergame for Shoulder Rehabilitation and Physical Fitness. : IEEE; 2014 Presented at: 2014 IEEE Games Media Entertainment; 2014 Oct 22-24; Toronto. [doi: 10.1109/GEM.2014.7224735]
23. Yin ZX, Xu HM. A wearable rehabilitation game controller using IMU sensor. : IEEE; 2018 Presented at: 2018 IEEE International Conference on Applied System Invention; 2018 Apr 13-17; Chiba p. 1060-1062. [doi: 10.1109/icasi.2018.8394459]

24. Wiederhold BK, Wiederhold MD. Evaluation of virtual reality therapy in augmenting the physical and cognitive rehabilitation of war veterans. Int J Disabil Human Dev 2006;5(3):211-215. [doi: 10.1515/ijdhd.2006.5.3.211]
25. Budziszewski P. A Low Cost Virtual Reality System for Rehabilitation of Upper Limb. In: VAMR 2013: Virtual, Augmented and Mixed Reality. Systems and Applications. Berlin, Heidelberg: Springer; 2013 Presented at: VAMR 2013; 2013 July 21-26; Las Vegas p. 32-39. [doi: 10.1007/978-3-642-39420-1_4]
26. Wilson RD, Chae J. Hemiplegic Shoulder Pain. Phys Med Rehabil Clin N Am 2015 Nov;26(4):641-655. [doi: 10.1016/j.pmr.2015.06.007] [Medline: 26522903]
27. Raghavan P. Upper Limb Motor Impairment After Stroke. Phys Med Rehabil Clin N Am 2015 Nov;26(4):599-610 [FREE Full text] [doi: 10.1016/j.pmr.2015.06.008] [Medline: 26522900]
28. Fitzgibbons P, Medvedev G. Functional and Clinical Outcomes of Upper Extremity Amputation. J Am Acad Orthop Surg 2015 Dec;23(12):751-760. [doi: 10.5435/JAAOS-D-14-00302] [Medline: 26527583]
29. Rahman M, Kankanhalli A, Wadhwa B, Hua Y, Kei C, Hoon LJ. GEAR: A Mobile Game-Assisted Rehabilitation System. : IEEE; 2016 Presented at: 2016 IEEE International Conference on Healthcare Informatics; 2016 Oct 4-7; Chicago p. 4-7. [doi: 10.1109/ichi.2016.92]
30. Lucchesi I, Lorussi F, Bellizzi M, Carbonaro N, Casarosa S, Trotta L. Daily Life Self-management and Self-treatment of Musculoskeletal Disorders Through SHOULPHY. In: MobiHealth 2017: Wireless Mobile Communication and Healthcare. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 247. Cham: Springer; 2018 Presented at: MobiHealth 2017; 2017 November 14-15; Vienna p. 233-241. [doi: 10.1007/978-3-319-98551-0_27]
31. Chang CM, Chang YC, Hsiao BY. The design of a shoulder rehabilitation game system. London: IET; 2010 Presented at: International Conference on Frontier Computing. Theory, Technologies and Applications; 2010 Aug 4-6; Taichung. [doi: 10.1049/cp.2010.0553]
32. Symeonidis I, Kavallieratou E. Development and Assessment of a Physiotherapy System Based on Serious Games. In: Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing. Cham: Springer; 2016 Presented at: MEDICON 2016; 2016 March 31st-April 2nd; Paphos p. 592-595. [doi: 10.1007/978-3-319-32703-7_115]
33. Rahman M, Wadhwa B, Kankanhalli A, Hua YC, Kei CK. GEAR analytics: A clinician dashboard for a mobile game assisted rehabilitation system. 2016 Presented at: 4th International Conference on User Science and Engineering; 2016 Aug 23-25; Maleka. [doi: 10.1109/iuser.2016.7857959]
34. The Boston Consulting Group. Realizing the Promise of Disease Management: Payer Trends and Opportunities in the United States. Boston: The Boston Consulting Group; 2006. URL: https://www.bcg.com/documents/file14744.pdf [accessed 2020-04-06]
35. The Boston Consulting Group. The promise of disease management. Boston: The Boston Consulting Group; 1995.
36. Mangal NK, Pal S, Khosla A. Frozen shoulder rehabilitation using microsoft kinect. : IEEE; 2017 Presented at: 2017 International Conference on Innovations in Green Energy and Healthcare Technologies; 2017 Mar 16-18; Coimbatore. [doi: 10.1109/igeht.2017.8094043]
37. Vogt K, Pirrò D, Kobenz I, Höldrich R, Eckel G. PhysioSonic - Evaluated Movement Sonification as Auditory Feedback in Physiotherapy. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen K, editors. CMMR 2009, ICAD 2009: Auditory Display. Berlin, Heidelberg: Springer; 2010:103-120.
38. Huang SY, Yu JP, Wang YK, Liu JW. Designing an exergaming system for exercise bikes using kinect sensors and Google Earth. Multimed Tools Appl 2016 Jun 2;76(10):12281-12314. [doi: 10.1007/s11042-016-3641-6]
39. Pekyavas NO, Ergun N. Comparison of virtual reality exergaming and home exercise programs in patients with subacromial impingement syndrome and scapular dyskinesis: Short term effect. Acta Orthop Traumatol Turc 2017 May;51(3):238-242 [FREE Full text] [doi: 10.1016/j.aott.2017.03.008] [Medline: 28446376]
40. Rizzo JR, Thai P, Li EJ, Tung T, Hudson TE, Herrera J, et al. Structured Wii protocol for rehabilitation of shoulder impingement syndrome: A pilot study. Ann Phys Rehabil Med 2017 Nov;60(6):363-370 [FREE Full text] [doi: 10.1016/j.rehab.2016.10.004] [Medline: 28330586]
41. Chang CM, Chang YC, Chang HY, Chou LW. An interactive game-based shoulder wheel system for rehabilitation. Patient Prefer Adherence 2012;6:821-828 [FREE Full text] [doi: 10.2147/PPA.S37190] [Medline: 23226005]
42. Pinto JF, Carvalho HR, Chambel GRR, Ramiro J, Goncalves A. Adaptive gameplay and difficulty adjustment in a gamified upper-limb rehabilitation. 2018 Presented at: 6th International Conference on Serious Games and Applications for Health; 2018 May 16-18; Vienna. [doi: 10.1109/segah.2018.8401363]
43. Arif A, Maulidevi NU, Dharma D, Alimansyah MR, Prabowo T. An Interactive Kinect-Based Game Development for Shoulder Injury Rehabilitation. 2018 Presented at: 5th International Conference on Data and Software Engineering; 2018 Nov 7-8; Senggigi Beach. [doi: 10.1109/icodse.2018.8705844]
44. Chung CE, Chen CH. The App Game Interface Design for Frozen Shoulder Rehabilitation. In: Soares M, Falcão C, Ahram TZ, editors. Advances in Ergonomics Modeling, Usability & Special Populations. Cham: Springer; 2017:507-516.
45. Fikar P, Schoenauer C, Kaufmann H. The Sorcerer's Apprentice A serious game aiding rehabilitation in the context of Subacromial Impingement Syndrome. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; 2013 May 5-8; Venice. [doi: 10.4108/icst.pervasivehealth.2013.252224]

46. Fernandez-Cervantes V, Stroulia E, Castillo C, Oliva L, Gonzalez F. Serious rehabilitation games with Kinect. 2015 Presented at: 2015 IEEE Games Entertainment Media Conference; 2015 Oct 14-16; Toronto. [doi: 10.1109/gem.2015.7377254]
47. Powell V, Powell W. Therapy-led Design of Home-Based Virtual Rehabilitation. 2015 Presented at: IEEE 1st Workshop on Everyday Virtual Reality; 2015 Mar 23; Arles. [doi: 10.1109/wevr.2015.7151688]
48. Nava W, Mejia CAR, Uribe-Quevedo A. Prototype of a ShoulderElbow Occupational Health Care Exergame. In: Stephanidis C, editor. HCI International 2015 - Posters' Extended Abstracts. Cham: Springer; 2015:467-472.
49. Du YC, Shih CB, Fan SC, Lin HT, Chen PJ. An IMU-compensated skeletal tracking system using Kinect for the upper limb. Microsyst Technol 2018 Feb 13;24(10):4317-4327. [doi: 10.1007/s00542-018-3769-6]
50. Huang MC, Lee SH, Yeh SC, Chan RC, Rizzo A, Xu W, et al. Intelligent Frozen Shoulder Rehabilitation. IEEE Intell. Syst 2014 May;29(3):22-28. [doi: 10.1109/mis.2014.35]
51. Muñoz GF, Casero JG, Cárdenas RAM. Usability Study of a Kinect-Based Rehabilitation Tool for the Upper Limbs. In: Rocha Á, Adeli H, Reis L, Costanzo S, editors. New Knowledge in Information Systems and Technologies. Cham: Springer; 2019:755-763.
52. Da Gama A, Chaves T, Figueiredo L, Teichrieb V. Guidance and Movement Correction Based on Therapeutics Movements for Motor Rehabilitation Support Systems. 2012 Presented at: 14th Symposium on Virtual and Augmented Reality; 2012 May 28-31; Rio Janiero. [doi: 10.1109/svr.2012.15]
53. Yeh SC, Lee SH, Fan YC. The Development of Interactive Shoulder Joint Rehabilitation System Using Virtual Reality in Association with Motion-Sensing Technology. In: Huang YM, Chao HC, Deng DJ, Park J, editors. Advanced Technologies, Embedded and Multimedia for Human-centric Computing. Dordrecht: Springer; 2014:1073-1082.
54. García-Martínez S, Orihuela-Espina F, Sucar LE, Moran AL, Hernández-Franco J. A design framework for arcade-type games for the upper-limb rehabilitation. 2015 Presented at: 2015 International Conference on Virtual Rehabilitation; 2015 June 9-12; Valencia. [doi: 10.1109/icvr.2015.7358573]
55. Haber P. Leitfaden zur medizinischen Trainingsberatung - Rehabilitation bis Leistungssport. Berlin, Heidelberg: Springer-Verlag; 2005.
56. Gears D, Braun K. Gamification in Business: Designing Motivating Solutions to Problem Situations. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2013 Apr 27-May 2; Paris URL: http://gamification-research.org/wp-content/uploads/2013/03/Gears_Braun.pdf
57. Reeves B, Cummings JJ, Anderson D. Leveraging the engagement of games to change energy behavior. 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2011 May 7-12; Vancouver. [doi: 10.1109/CTS.2012.6261074]
58. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
59. Quevedo WX, Ortiz JS, Velasco PM, Sánchez JS, Álvarez M, Rivas D. Assistance System for Rehabilitation and Valuation of Motor Skills. In: De Paolis L, Bourdot P, Mongelli A, editors. Augmented Reality, Virtual Reality, and Computer Graphics. Cham: Springer; 2017:166-174.
60. Nussbaum R, Kelly C, Quinby E, Mac A, Parmanto B, Dicianno BE. Systematic Review of Mobile Health Applications in Rehabilitation. Arch Phys Med Rehabil 2019 Jan;100(1):115-127. [doi: 10.1016/j.apmr.2018.07.439] [Medline: 30171827]
61. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business. Pennsylvania: Wharton Digital Press; 2010.

Abbreviations

PICO: population, intervention, comparison, outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by R Haghighi Osgouei, M Rosly; comments to author 02.07.20; revised version received 07.07.20; accepted 22.07.20; published 25.08.20.

Please cite as:
Steiner B, Elgert L, Saalfeld B, Wolf KH
Gamification in Rehabilitation of Patients With Musculoskeletal Diseases of the Shoulder: Scoping Review
JMIR Serious Games 2020;8(3):e19914
URL: http://games.jmir.org/2020/3/e19914/
doi:10.2196/19914
PMID:32840488

Original Paper

Feasibility of a Commercially Available Virtual Reality System to Achieve Exercise Guidelines in Youth With Spina Bifida: Mixed Methods Case Study

Byron Lai[1*], PhD; Drew Davis[1*], MD; Mai Narasaki-Jara[2*], EdD; Betsy Hopson[3*], MSHA; Danielle Powell[4*], MD; Marissa Gowey[5*], PhD; Brandon G Rocque[3*], MD, MS; James H Rimmer[6*], PhD

[1]Division of Pediatric Rehabilitation Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Department of Kinesiology, California Polytechnic State University Pomona, Pomona, CA, United States

[3]Division of Pediatric Neurosurgery, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Department of Physical Medicine and Rehabilitation, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[5]Division of Pediatric Gastroenterology, Hepatology, and Nutrition, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[6]Dean's Office, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[*]all authors contributed equally

Corresponding Author:
Byron Lai, PhD
Division of Pediatric Rehabilitation Medicine
School of Medicine
University of Alabama at Birmingham
1720 University Blvd
Birmingham, AL, 35294
United States
Phone: 1 2052001724
Email: byronlai@uab.edu

Abstract

Background: Access to physical activity among youth with spina bifida (SB) is much lower than it is for children without disability. Enjoyable home-based exercise programs are greatly needed.

Objective: Our objective is to examine the feasibility of a virtual reality (VR) active video gaming system (ie, bundle of consumer-available equipment) to meet US physical activity guidelines in two youth with SB.

Methods: Two youth with SB—a 12-year-old female and a 13-year-old male; both full-time wheelchair users—participated in a brief, 4-week exercise program using a popular VR head-mounted display: Oculus Quest (Facebook Technologies). The system included a Polar H10 (Polar Canada) Bluetooth heart rate monitor, a no-cost mobile phone app (VR Health Exercise Tracker [Virtual Reality Institute of Health and Exercise]), and 13 games. The intervention protocol was conducted entirely in the homes of the participants due to the coronavirus disease 2019 (COVID-19) pandemic. The VR system was shipped to participants and they were instructed to do their best to complete 60 minutes of moderate-intensity VR exercise per day. Exercise duration, intensity, and calories expended were objectively monitored and recorded during exercise using the heart rate monitor and a mobile app. Fatigue and depression were measured via self-report questionnaires at pre- and postintervention. Participants underwent a semistructured interview with research staff at postintervention.

Results: Across the intervention period, the total average minutes of all exercise performed each week for participants 1 and 2 were 281 (SD 93) and 262 (SD 55) minutes, respectively. The total average minutes of moderate-intensity exercise performed per week for participants 1 and 2 were 184 (SD 103) (184/281, 65.4%) and 215 (SD 90) (215/262, 82.1%) minutes, respectively. One participant had a reduction in their depression score, using the Quality of Life in Neurological Disorders (Neuro-QoL) test, from baseline to postintervention, but no other changes were observed for fatigue and depression scores. Participants reported that the amount of exercise they completed was far higher than what was objectively recorded, due to usability issues with the chest-worn heart rate monitor. Participants noted that they were motivated to exercise due to the enjoyment of the games and VR headset as well as support from a caregiver.

Conclusions: This study demonstrated that two youth with SB who used wheelchairs could use a VR system to independently and safely achieve exercise guidelines at home. Study findings identified a promising protocol for promoting exercise in this population and this warrants further examination in future studies with larger samples.

(JMIR Serious Games 2020;8(3):e20667) doi:10.2196/20667

KEYWORDS

physical activity; active video gaming; exergaming; disability; Oculus Quest

Introduction

Spina bifida (SB) is a common birth defect that results in permanent mobility disability and affects ~170,000 people in the United States [1]. Approximately 37% percent of people with SB use a wheelchair or other assistive device as a primary means of mobility [2]. The importance of addressing obesity and physical inactivity (ie, insufficient levels of physical activity or its subset, exercise, to achieve health benefits) in this population cannot be understated. Approximately 37% of people with SB are obese [3], and they have higher fat mass (159%), lower cardiorespiratory endurance (32%-54% peak oxygen volume [VO_2]), and lower muscular strength (58%-98%) compared to peers without SB [4]. Over time, these factors result in rapid physical deconditioning, placing them at risk for metabolic syndrome, cardiovascular disease, and type 2 diabetes [5]. There is a pressing need to identify effective approaches for living well with SB.

The transition from childhood to adulthood is a critical threshold for addressing the health needs of individuals with SB. Regular exercise is one of the most important health behaviors for reducing the risks associated with physical inactivity. Health professionals will typically instruct adolescents to try and achieve the exercise guidelines for children of 60 minutes per day of moderate-to-vigorous-intensity activity [6]. However, these recommendations are often unattainable for people who are physically inactive and there is evidence that adolescents with disabilities benefit from far lower volumes of physical activity [7]. Thus, health professionals will more commonly prescribe the adult exercise guidelines of 150 minutes of moderate-to-vigorous-intensity activity per week [6]. Youth who exercise regularly are more likely to live healthy, active lifestyles as adults and reduce their risk of chronic conditions.

For people with SB, several core concerns prompted the need for this pilot study:

1. Obesity rates for children with SB become alarmingly higher in adulthood [3].
2. Accessible and usable exercise opportunities within the community are scarce for youth with SB as they transition into adulthood [8].
3. Feelings of isolation and depression increase exponentially when transitioning into adulthood [9].
4. Social relationships, vocations or meaningful hobbies, and reduced dependence on caregivers are core developmental tasks during adolescence [10].
5. Wheelchair reliance increases as children with SB enter adolescence, which is partially due to excessive weight gain [2].

As reported in a systematic review [11], aerobic and strength interventions for youth with SB typically last 11 weeks, include two to three sessions per week, and total ~60 minutes of exercise per week. These encouraging, preliminary studies demonstrate potentially efficacious methods for improving aerobic capacity (ie, three studies, effect size range 0.74-1.40) and strength (ie, two studies, effect size range 0-0.59). However, these interventions included supervision from a therapist or research staff. To date, there is no program that can promote self-regulated exercise behavior among youth with SB, particularly individuals with mobility limitations who may experience difficulty with transportation to an on-site facility. Further research efforts are needed to identify accessible, enjoyable, and sustainable programs that can be easily engaged in by youth with SB.

Home-based exercise programs that deliver and monitor exercise prescriptions using telecommunication technology (ie, telehealth) are a desirable approach for people with disabilities who may not have convenient access to other means of exercise. The advantages of a telehealth approach over usual care include the following: increased social support, participant adherence, quality of care, cost-effectiveness, access to services, and, most notably, reduced trainer burden to allow easier dissemination of services [12]. Barriers to exercise can include lack of nearby accessible facilities, usable equipment, knowledgeable staff, and transportation [13,14]. For these reasons, exercise through telehealth is rapidly becoming the new norm during the coronavirus disease 2019 (COVID-19) pandemic.

Active video gaming (AVG) is an enjoyable method of home exercise that has the potential to engage youth with SB in long-term exercise behavior. AVG can be performed at a moderate intensity of exercise [15], making it a potentially effective method for improving health, function, and body composition. Virtual reality (VR)—computer and sensor technology that is used to create a simulated environment or experience—is an emerging area of AVG technology. In rehabilitation, VR technology is being used as an enjoyable method for managing pain and improving motor and executive function, fitness, movement quality, and spatial orientation and mobility [15-18]. However, only a handful of studies have investigated the influence of VR technology on promoting exercise behavior outside of a formal rehabilitation setting among youth with disabilities [15,16,19,20]. Furthermore, home-based studies have primarily incorporated less immersive forms of VR, including the Nintendo Wii and Xbox Kinect [15], but these technologies have been discontinued by game manufacturers.

Head-mounted display (HMD) technology is the latest trend in AVG. In May 2019, one of the largest investments in VR gaming by Facebook led to a pivotal advancement in making HMD more ubiquitous for consumers: the development of the Oculus Quest (Facebook Technologies). The Quest is the first VR headset of high visual quality (ie, up to 72 frames per second) that does not require any plug-in to a costly desktop gaming computer or game console. Higher frame rates provide a smoother visual display (aka, immersion) within a virtual environment [21], whereas lower frame rates can induce motion sickness and nausea [22]. The Quest has access to the latest top-rated VR fitness games and has online multiplayer capability that comes with free-to-play multiplayer cooperative and competitive recreational games. This makes the Quest the first off-the-shelf, all-in-one headset that allows health professionals to prescribe a high-quality immersive, enjoyable, and socially connected VR experience in the comfort of one's home.

While there are a few studies that have demonstrated successful exercise responses in youth with SB [23,24], these studies have used nonstandalone HMDs: HTC Vibe and Oculus Rift. These devices had to be plugged in to a computer gaming system, which was stationed on-site at a school or rehabilitation clinic. To our knowledge, no study has demonstrated whether the Quest can be used for exercise training among children with disabilities, including SB. Moreover, we are not aware of any studies that promoted unsupervised exercise behavior that met or exceeded the duration component of the US exercise guidelines in youth with SB.

The case study in this paper tested the feasibility of using the Oculus Quest VR system to achieve exercise guidelines among two youth with SB who used a wheelchair as their primary means of mobility. Feasibility was evaluated via two *usability* metrics: *effectiveness* and *usefulness* [25]. The aims of this study were as follows:

1. Aim 1 (*effectiveness*): compare the minutes of moderate-intensity exercise achieved each week with exercise guidelines for children (60 min/day) and adults (150 min/week).
2. Aim 2 (*usefulness*): describe participants' perceptions of factors that affected their ability to use the system to achieve the exercise guidelines.

The ancillary aim was to describe the potential treatment effects of the VR program on fatigue and depression.

Methods

Overview

The design was a mini-ethnographic case study for a convenience sample of two people [26,27]. A mini-ethnographic case design, also known as focused ethnography, is used for retrospectively developing a rich understanding of an individual or group response to a program or study [26]. This exploratory approach is smaller in scope and generally shorter in duration than a full-scale ethnographic approach, which embeds a researcher within a setting for a prolonged period of time to examine the lived experience through more pattern-focused analytical techniques (eg, grounded theory or thematic analysis) [26]. Ethnography is ideal for investigating the potential of innovative products or programs that are commonly used in medical and marketing research [26,28]. This design incorporated a nested mixed methods approach—QUANTITATIVE → qualitative [29]—for a comprehensive evaluation of feasibility. Specifically, the design included a qualitative postintervention interview within a primarily quantitative, feasibility preintervention-to-postintervention design. Quantitative and qualitative methods were combined within the Discussion section of this paper to provide an *expanded* evaluation of usability.

Recruitment

This study purposefully selected one male and one female youth from the Spina Bifida Clinic at the Children's Hospital of Alabama. Inclusion criteria were as follows: (1) less than 18 years of age, (2) used a wheelchair as a primary means of mobility, (3) physically inactive: in the previous 2 weeks, self-reported less than 150 minutes per week of moderate-to-vigorous-intensity exercise and no participation in a structured exercise program or regimen, (4) diagnosis of SB, (5) access to a mobile device: computer tablet or smartphone, and (6) Wi-Fi internet access in the home. Exclusion criteria were as follows: (1) owned a VR headset device or (2) had a health condition that prevented participation in moderate-intensity exercise. This study was conducted in accordance with case study guidelines set by the Institutional Review Board for Human Use at the University of Alabama at Birmingham (UAB). Written informed consent was obtained from participants prior to participation.

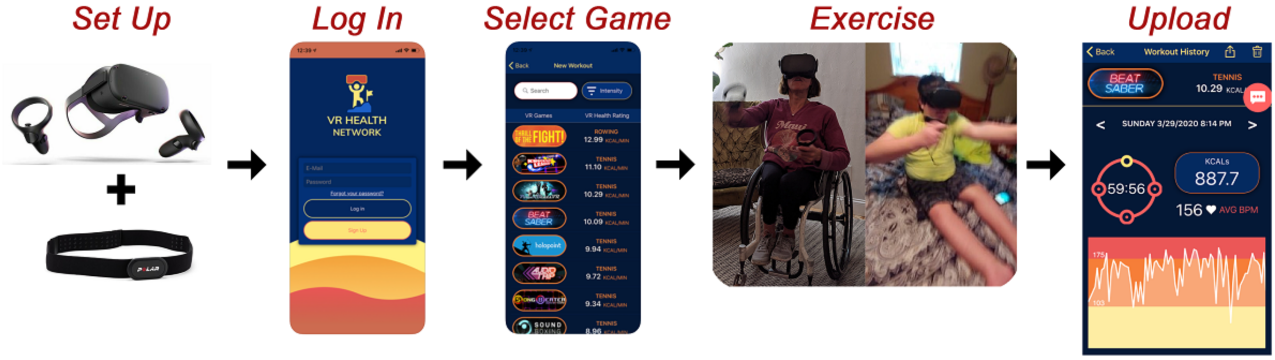
Intervention Protocol

Participants were provided with a VR unit and asked to use the VR system to complete 60 minutes of moderate-intensity exercise per day, a guideline set by the US Department of Health and Human Services [6].

VR System

The VR system consisted of the following components (see Figure 1):

Figure 1. Exercise session procedures of the virtual reality (VR) system.



1. An Oculus Quest headset (64 GB) with two hand-held controllers.
2. A Polar H10 (Polar Canada) Bluetooth, chest strap, heart rate monitor.
3. A no-cost mobile phone app—VR Health Exercise Tracker (Virtual Reality Institute of Health and Exercise)—that records exercise duration and intensity.
4. A total of 13 games installed on the headset.

Prior to the start of the intervention, participants and one parent had a 30-minute videoconference call with the telecoach. The purpose of the call was to make sure they received the equipment and understood how to set up the system, reviewed the study procedures (ie, study objectives, the VR system, exercise prescription, and the postintervention interview), and briefly summarized the benefits of exercise. In accordance with the Supportive Accountability Theory [30], the goal of the brief communications—primarily the initial coaching call—was to develop a friendly relationship and bond with the participant and establish a social presence that provided a sense of accountability to enhance participant motivation and adherence to the study procedures.

In addition to the call, participants were provided with two pages of printed instructions. The instructions were presented in a bullet-list format that guided participants through the following setup procedures: (1) log in to the headset with an Oculus account that was created by the researcher; no identifying information from the participant was inputted or connected to an Oculus or Facebook account, (2) download the VR Health Exercise Tracker app on their phone, and (3) sync the Polar heart rate monitor to the VR Health Exercise Tracker app. The instructions also guided participants through the exercise session procedures: (1) log in to the VR Health Exercise Tracker app, (2) select a game, (3) turn on and equip the headset, (4) exercise with the selected game, and (5) after exercise, review and upload the exercise data into a cloud-based server.

During exercise, the VR Health Exercise Tracker app automatically recorded exercise data via Bluetooth from the Polar monitor. The Polar electrocardiographic chest strap monitors, including the H10, have over two decades of published studies demonstrating strong psychometric properties to support their use for measuring heart rate among both youth and adults [31-33]. VR Health Exercise Tracker records and displays in real time the following data: exercise intensity (heart rate), calories, and exercise duration (minutes). After each session, participants reviewed their exercise data and uploaded the data into a private folder within a cloud-based server (Google Drive). Based on the Tanaka equation [34], participant 1 (pseudonym: Johnny) was instructed to keep his exercise heart rate between 119 and 159 beats per minute; participant 2 (pseudonym: Sapphire) was instructed to keep her exercise heart rate between 120 and 160 beats per minute.

A telecoach (BL) reviewed and recorded exercise data obtained from the cloud-based server on a weekly basis. At the end of each week, the telecoach emailed the participants a progress report with graphs that displayed the total minutes and moderate-to-vigorous-intensity minutes of exercise completed. The reports included typed comments with positive verbal reinforcement of their exercise behavior, along with reminders to try their best to achieve the exercise prescription goals.

Intervention Games

The Quest headset that was sent to participants was preinstalled with 13 games. The games required dynamic, speed-driven movements of the arms that could be played in a seated position. The games included rhythmic movements to music as well as sport and recreation activities that elicited high energy expenditure (eg, dancing, boxing, and tennis), as determined by the Virtual Reality Institute of Health and Exercise [35]. Nine of the 13 fitness games were chosen by the telecoach because they could be performed while standing or sitting and had a high development quality. These core games were Beat Saber (movement to music), The Climb (mountain climbing), Racket Fury (table tennis), Thrill of the Fight (boxing), Sports Scramble (various sport and recreation activities), Half + Half, Rec Room, Dance Central, and a local multiplayer game that could be played with the family (Acron: Attack of the Squirrels!). Two of these games, Rec Room and Half + Half, were online group-based multiplayer games that could be downloaded at no cost from the Oculus Store. Acron: Attack of the Squirrels! is a cooperative and competitive game that matches up players (ie, squirrels) who use their personal mobile phones or tablet devices against the VR headset user (ie, tree); there is a maximum of 8 players. These three multiplayer games did not have a single-player game mode. Sports Scramble supported single-player and online multiplayer gameplay. Participants were allowed to choose the remaining four fitness games that they wanted installed on their headset. Chosen games included the following: Creed: Rise to Glory and BOXVR (two boxing games), Cloudlands 2 (golf), Racket: Nx (table tennis), AUDICA (movement to music), Synth Riders, and Moss.

Participants were encouraged to test the suitability of each game for wheelchair exercise.

Measures

The data collection was originally planned to be conducted on-site at a local fitness facility that had a human performance laboratory. In response to the societal impact of COVID-19, all data were collected remotely. Participants archived their data from each exercise session on the cloud-based server. Participants completed the study questionnaires before starting the 4-week intervention and a second time after they completed the 4-week intervention. Research staff sent the questionnaires to participants via email. Participants uploaded the completed questionnaires to the cloud-based server.

Fatigue and depression were measured via self-report, short-form questionnaires from the Neuro-QoL (Quality of Life in Neurological Disorders) Pediatric archive [36]. Each questionnaire had eight items that could be scored on a scale ranging from 1 (none of the time) to 5 (all of the time). Higher scores reflect higher levels of fatigue and depression. Neuro-QoL tests were systematically developed by a multisite collaboration and psychometrically evaluated among a large population-based sample [37]; they are currently being tested for further validation among more specific groups of people with neurological conditions (eg, traumatic brain injury, epilepsy, and muscular dystrophy) [38,39], albeit not yet among people with SB. Additional Neuro-QoL details and an extensive list of validation studies can be found elsewhere [36].

Postintervention Interview

Participants underwent a semistructured interview with the researcher (BL) at the end of the program. The interview script is provided in Textbox 1. The interview included 10 overarching questions with several follow-up questions. The questions probed participants' overall views of the program, likes and dislikes, experiences using the VR equipment for exercise during the program, barriers and facilitators to meeting the exercise prescription goals, preferences for the fitness games, and recommendations to enhance the protocol for a future trial. The interviews were conducted through videoconference. The participants were accompanied by their caregivers. The principal investigator (BL) had a background in adapted physical activity, along with several years of experience conducting qualitative interviews related to exercise and disability. Each interview lasted approximately 2 hours.

Textbox 1. Interview script.

Interview Briefing:

“Good [morning, afternoon, evening]. This interview will include 10 questions, along with several follow-up questions. The questions will explore your perceptions of using the Oculus Quest headset to achieve exercise recommendations. This interview will be recorded so that we can later type and analyze the data for publication as a research study. This study is completely confidential and we will not include your name or any other identifying information. To keep it confidential, please don’t say any names of people or places. Your participation in the study will not affect any services or relationships you have with people at the University or Hospital.”

“Before we start, please give me a fun and fake name to call you so that I do not say your real name during the interview.”

Participant Pseudonym:_________________

*Audio Recording Started (TURN ON AUDIO RECORDING and go to next page):

“I have now started the audio recording. Once again, your participation in this study is totally voluntary and your verbal consent over the phone now indicates that we have gone over the study details and that we have permission from you and one of your parents to participate in this phone call interview.”

“Just so I can get a picture of you, I’m going to ask you a few questions”:

1. How old are you?

2. What ethnicity are you?

3. How tall are you?

4. How much do you weigh?

5. Do you wear orthotics or use ambulatory aides? If so, what kind?

6. Do you use a wheelchair? If so, what kind?

7. Generally, do you live in a rural area or urban city area?

“Okay, before we get started, please be open and honest with your responses. Please say the first things that come to your head: what you think and what you feel. There are no right or wrong answers. Please just tell your story. Even though we are on the telephone, don’t be afraid to keep talking. I will try to be listening more than talking during the interview.”

**Interview Questions* (Prompts: “Could you tell me more about that?,” “How did that make you feel?,” “Do you have anything else you’d like to add about that?”):

1. As an icebreaker, tell me about yourself. What do you like to do for fun or in your free time?

2. Okay, please take me through what a typical week/day is for you.

3. What does exercise mean to you? (images, words, activities, anyone ever talk to you about it?, advice)

“Okay, this next question will be a bit different. For this question, we will discuss your story about your experiences with the virtual reality headset, the Oculus Quest. We will talk about your thoughts of exercise starting from when you first started the program all the way to the end of the month.”

4. Okay, so let’s start. Before you started the program, like when I first called you, how did you feel about starting an exercise program using the virtual reality headset?

Follow-up Questions:

i. How confident were you that you could exercise with the technology?

ii. Before you started the program, did you exercise? (how often?)

iii. How did you feel about exercise?

iv. Were you feeling motivated or energized to do the exercises?

5. Okay, so in the first week of the program (show data). What did you think about exercising with the headset?

Follow-up Questions:

i. How confident were you that you could exercise with the technology?

ii. Were there any problems you had with using the technology for exercise?

iii. Were there any things you really liked?

iv. Were there any things you really didn’t like?

v. Was it difficult to do 60 minutes of exercise each day?

vi. Was it difficult to do 60 minutes of MODERATE-intensity exercise each day?

vii. Were you feeling motivated or energized to do the exercises?

viii. Were there any issues with the technology?

ix. Were there any things you think we could have done in this first week to make the experience better?

6. Okay, so now let's talk about the second week of the program (show data). What did you think about exercising with the headset during the second week?

Same Follow-up Questions as 5

7. Okay, so now let's talk about the third week of the program (show data). What did you think about exercising with the headset during the third week?

Same Follow-up Questions as 5

8. Okay, so now let's talk about the fourth week of the program (show data). What did you think about exercising with the headset during this last week?

Same Follow-up Questions as 5

9. Okay, next question; we are planning to give the same program you went through to other kids with spina bifida around your age. Are there any thoughts you have that could make the program better for other kids?

10. Okay, video game time! What was your favorite game to play?

Follow-up Questions:

i. What did you like about each game? Did you find it easy to get your heart rate to the moderate zone from that game?

ii. For each game you mentioned, what are some tips or strategies you can recommend for other kids? (Go through each game they can recall).

iii. What was your least favorite game? (Follow up with why).

11. Do you have anything else you would like to add?

Analysis

All quantitative outcomes (ie, objectively recorded exercise data and self-reported questionnaire data) were descriptively reported. To address Aim 1, the average minutes of moderate-to-vigorous-intensity exercise performed each week were compared with both the youth and adult guidelines: ≥420 minutes of moderate-to-vigorous-intensity exercise per week (ie, >60 minutes per day) was classified as *excellent* (youth guidelines); ≥150 minutes of moderate-intensity exercise per week was classified as *sufficiently active* (adult guidelines); and <150 minutes of moderate-intensity exercise per week was classified as *insufficiently active* (adult guidelines). Neuro-QoL depression and fatigue scores at baseline and postintervention were summed and converted to T scores [40].

The qualitative component utilized an explanatory narrative approach [41] to explain the potential usefulness of the program and expand interpretations of the quantitative data. Interview data were transcribed and double-checked for accuracy. The results were reported in narrative format, underpinned by an interpretivism paradigm. The specific philosophical assumptions that underpinned the qualitative study methods were critical realism (ontological perspective) [42] and interpretivism (epistemological perspective) [43]. In other words, the researcher team acknowledged that youth with SB perceived a reality when reporting their responses, and the recollection of this reality or experience was recreated by the interaction between the youth and the interviewer, as well as the interpretation of the data.

Results

Overview

Characteristics of the two participants are presented in Table 1. The quantitative exercise data are reported below, followed by an explanatory narrative of each case. There were no adverse events reported by participants throughout the program. The participants completed the program in April 2020, a period when states within the United States were under a *social distancing* and *shelter-in-place* order.

Table 1. Case characteristics.

Characteristic	Case 1: Johnny	Case 2: Sapphire
Age (years)	13	12
Sex	Male	Female
Race	Caucasian	Caucasian
Height (feet, inches)	5, 3	5, 0
Weight (pounds)	174	132
Body mass index (kg/m^2)	30.8	25.8

Quantitative Exercise Data

The total average (SD) minutes of exercise performed each week for case 1 (Johnny) and case 2 (Sapphire) were 281 (SD 93) and 262 (SD 55) minutes, respectively. The total average minutes of moderate-to-vigorous-intensity exercise per week for cases 1 and 2 were 184 (SD 103) and 215 (SD 90) minutes, respectively. Accordingly, both case participants were classified as *sufficiently active*. These values were below the exercise guidelines for children, but exceeded exercise guidelines for adults (see Figure 2). On average, participants spent 71.1% (SD 12%) of their time exercising at a moderate intensity. Weekly exercise data are reported in Table 2.

Figure 2. Moderate-intensity exercise minutes completed by each participant. P1: participant 1, Johnny; P2: participant 2, Sapphire.

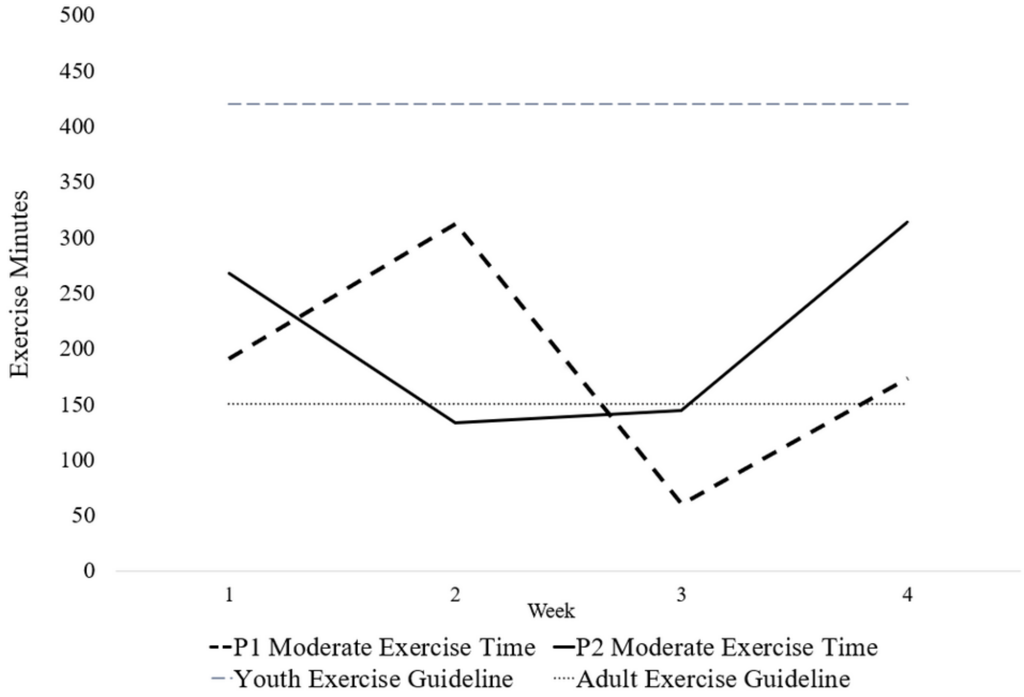


Table 2. Weekly exercise data.

Measure	Participant 1: Johnny	Participant 2: Sapphire
Moderate-intensity exercise (minutes), n/N (%)		
Week 1	191/364 (52.5)	268/304 (88.2)
Week 2	312/312 (100)	133/212 (62.7)
Week 3	60/147 (40.8)	144/216 (66.7)
Week 4	173/299 (57.9)	314/314 (100)
Number of days exercised	24	15
Heart rate per day of exercise (beats per minute), mean (SD)	127 (17)	127 (15)
Caloric expenditure per day of exercise (kcal), mean (SD)	496 (215)	441 (145)
Total number of exercise sessions	29	29
Total number of moderate-intensity exercise sessions	21	25

Both participants reported average levels of depression at baseline (case 1: Neuro-QoL score of 12, T score=47.9; case 2: Neuro-QoL score of 13, T score=48.8). Case 1 reported an improvement in depression from baseline to postintervention (change in T score of 11.4) that surpassed a minimal detectable change of 9.6 at case level [40]. Case 2 reported no change in depression score at postintervention. Fatigue scores at baseline were low and did not change at postintervention (case 1: Neuro-QoL baseline score of 10, T score=36.5, and postintervention score of 12, T score=39.5; case 2: baseline Neuro-QoL score of 15, T score=42.8, and postintervention Neuro-QoL score of 14, T score=41.8).

Explanatory Qualitative Cases

Case Summary

The case summary is shown in the items below:

1. Participants engaged in no exercise prior to the intervention.
2. Before COVID-19, participants engaged in no exercise outside of a school setting, and they were not included in many physical education activities.

3. Participants were hesitant to join the study but became confident in their ability to exercise throughout the program.
4. Participants reported a high level of enjoyment and immersion with the VR headset and games, which motivated them to exercise.
5. Caregivers facilitated exercise behavior and helped participants adhere to the monitoring procedures.
6. Participants reported that they engaged in far more exercise than what was recorded objectively, due to issues with the chest-worn heart rate monitor.
7. Three preferences for active games were identified: opponent-driven fighting games, multiplayer games, and rhythmic movement-to-music games.
8. Participants reported several minor usability issues: low battery life, headset and heart rate monitor discomfort, troubles with online multiplayer play, and some games requiring modifications for accessibility.

Case 1: Johnny

Overview

Johnny was a 13-year-old Caucasian male that lived in a suburban area in Alabama, USA. Johnny used a wheelchair as a primary means of mobility and used an ankle-foot orthosis (AFO) when walking short distances. Prior to the COVID-19-related *shelter-in-place* order and the study, Johnny engaged in no exercise behavior outside of school physical education classes. Watching and attending sporting events, particularly football, was one of Johnny's favorite hobbies. Yet, he was unable to engage in most sports and activities with his peers during physical education class. The activities in his class were not fully adapted or modified for a wheelchair user. Thus, Johnny spent much of his physical education time isolated from participation on the sidelines. Once COVID-19 imposed quarantine restrictions, he engaged in no exercise behavior. Most of his free time was spent watching sports on television or playing nonactive video games.

At the start of the program, Johnny reported that he was not confident that he could complete 60 minutes of moderate-intensity exercise in a day. He felt intimidated by these instructions, and stated that the duration felt long and he could not remember the last time he completed 60 minutes of any exercise in one day. These feelings were maintained until he opened the box with the headset that was sent to his home. Immediately upon viewing the headset, Johnny was extremely excited by the quality of the technology and quickly set up the system and played the games. The first day he received the headset, he used it for over 3 hours. After using the VR headset for 1 week, he realized that exercising for 60 minutes a day at a moderate intensity could easily be achieved and did so in his first day of using the headset. He was very excited and proud of his ability to exercise with the headset and would consistently show his family some of the VR experiences and demonstrate his physical prowess during gameplay. His mother reported that she would have to limit his daily use of the headset so as not to interfere with mealtime and schoolwork. His mother also reported that he spent approximately three times as much time using the headset than what was recorded through the VR system: approximately 4 hours each day. Headset use mostly included active games, but some of this time included passive device use, such as a free-to-download VR roller coaster ride or "every once in a while" watching shows on Netflix. Johnny exercised with the headset in a wheelchair, chair, or bed or on the floor.

Facilitators

Johnny reported a high level of enjoyment while playing the fitness games, which was attributed to the immersion or "felt real" factor provided by the system and the fun games: "I loved it. All of it. I loved it." Johnny stated that his favorite games were those where he had to overcome an opponent—Creed: Rise to Glory and Thrill of the Fight—and the rhythmic movement-to-music game, Beat Saber. Johnny did not spend much of his time playing online multiplayer games, but one particularly enjoyable multiplayer game was Acron: Attack of the Squirrels! Johnny played this game with his entire family, which created some memorable experiences. Johnny reported that VR games were more enjoyable than typical console games and that he liked being able to play the game while moving his arms. His mother noted that he was so engaged with some of the games that his family members had to be careful around him: "My mom was sitting beside him and he knocked her glasses off and I was like, 'You have to watch out. He cannot see you' (laughs)." Prior to the VR program, Johnny noted that he spent 3-4 hours playing console games, such as Fortnite, or sports games each day, but he quit playing these games after using the VR headset. He stated that he preferred the VR games because "I like being able to play the game while I'm moving my arms." It is important to mention that Johnny reported that his motivation to exercise was primarily due to the enjoyment of the games and headset, but his mother highly motivated him to adhere to the exercise data collection protocol. Johnny's mother would consistently remind him to wear his heart rate monitor to record the data from each session. Johnny's mother was the person exporting the exercise data to a cloud-based server.

Notably, Johnny and his mother reported that he seemed to have lost some weight based on his facial features and some other areas of his body. Johnny's mother also mentioned a change in shirt size from extra-large to large from pre- to postintervention. Both Johnny and his mother were elated at being able to see an observable change in such a short period. Johnny was particularly thrilled that he was able to lose weight, which was an important goal of his, while doing activities that were so enjoyable that he was unaware he was exercising. He also reported that the calories expended after each session provided him with a sense of accomplishment. By accomplishing these feats and realizing that he was capable of performing a high volume of exercise, Johnny felt more motivated to engage in general exercise after the program. No objective measurement of weight was obtained, due to difficulties with obtaining body weight measurements while using a wheelchair. The last weight measurement that the family obtained was during a visit to their physician prior to the COVID-19 quarantine. Johnny's mother appreciated the opportunity to engage her son in the study, since the typical outdoor activities they engaged in during summer (eg, camp and parks) were unavailable due to the COVID-19 quarantine.

Barriers and Usability Issues

Johnny did much less exercise in weeks 2 and 3 than in weeks 1 and 4. Johnny and his mother noted that this was because he was traveling with his father in a conventional sleeper truck. Johnny brought the Quest with him while accompanying his father and would play in the truck while his father drove or before bed. Also, during week 3, Johnny did not feel well due to a flare-up of his allergies and this reduced his playing time. Johnny reported that he did not like to wear and equip the heart rate monitor before exercise. He noted that it felt uncomfortable and he would occasionally unfasten the strap during long bouts of exercise. For these reasons, much of the exercise he did was not recorded with the VR system. Johnny recommended that his heart rate be recorded from a wrist-worn device, such as his Apple Watch. Notably, Johnny mentioned that the battery life of the headset was too short in duration (~2 hours), which required him to consistently charge the headset after using it each day. Johnny reported that he tried the multiplayer games but did not engage in them regularly. He and his mother reported that there did not appear to be many other players when joining a game lobby. Unlike Xbox Live, they did not notice a big community of players. Instead, he spent more time playing local multiplayer games with his entire family. Family members played the games—Acron: Attack of the Squirrels! and Rec Room—using their mobile phones while Johnny played on the headset, and this provided them with several cherished opportunities to bond during the COVID-19 quarantine.

Case 2: Sapphire

Overview

Sapphire was a 12-year-old Caucasian female living in a suburban area in Alabama, USA. Sapphire used a wheelchair as a primary means of mobility and used an AFO when walking short distances. Similar to Johnny, Sapphire engaged in no exercise behavior outside of her school's physical education classes prior to the COVID-19 quarantine. At school, she was not included in most sports and activities with her peers during physical education class. Physical education classes included group activities, which Sapphire did not participate in: "I just watch the kids. I don't do anything." One day each week was devoted to exercises, and Sapphire felt more included within these activities. However, at a previous school she did have an adapted physical education teacher. The teacher developed a personalized program for Sapphire and modified activities to include her in the activities with peers. During the COVID-19 quarantine, Sapphire engaged in no exercise behavior and spent most of her free time watching television or using her phone. However, Sapphire and her mother reported that she had a high volume of schoolwork in the form of homework and virtual class sessions via videoconferencing. Since the family shared one laptop for work and school, they operated around a strict schedule. This meant that VR exercise had to be performed early in the morning before school or later in the evening.

Before Sapphire began the program, she was worried that she would not be able to complete the prescribed 60 minutes. The amount of time was perceived as too long and intimidating. Sapphire's mom encouraged her not to worry and to think more about playing than exercising. When the headset arrived at their home and Sapphire saw it, she was extremely excited and could not wait to try it out. She was surprised to realize that she completed 78 minutes of exercise in her first session. She said it was "very fun" and she could not wait to do it again. By the end of the study, it was very easy for her to complete 60 minutes of moderate exercise. She often wanted to play the VR games longer than prescribed and to get back on the headset later in the day. When exercising, Sapphire would sit in a wheelchair or chair.

Facilitators

Sapphire expressed a high level of enjoyment with the program. Similar to Johnny, Sapphire acknowledged that the virtual world "felt real" and was quick to report her excitement while playing certain games. Sapphire was so encapsulated by the virtual environment that she pleasantly described her experience and knowledge of four friends she made in a game, Dance Central, which the interviewer did not realize were in-game characters, not actual online players, until further clarification. However, the games that Sapphire preferred were slightly different than the opponent-driven fighting games preferred by Johnny. Sapphire preferred multiplayer games—AUDICA, Acron: Attack of the Squirrels!, Sports Scramble, and Rec Room—as well as rhythmic movement-to-music games: AUDICA and BOXVR. Sapphire specifically mentioned her enjoyment playing multiplayer games online, such as Rec Room, with other users and, most notably, at home with the family: Acron: Attack of the Squirrels! Sapphire consistently desired to play with her family and asked her father to play as soon as he returned home from work:

> *You ask him when he walks straight through the front door (laughs).* [Mother]
>
> *So? I want to play with him!* [Sapphire]

When asked how she was able to achieve so much exercise time, she stated that in the first week of the program she completed multiple 30-minute bouts as opposed to one 60-minute session. It became easier for her to complete 60-minute exercise bouts during the second week of the program. Another factor that motivated her to exercise was being able to view the calories she expended each day within the VR Health Exercise Tracker app.

Similar to Johnny, Sapphire's mom played a major role in the program. As stated by Sapphire's mom, "I helped her put the heart rate monitor on. I would get the VR off of her, charge it, hand it to her. She would put it on. I would help her secure the controllers to her wrist, so she doesn't accidentally throw them."

Sapphire's mom also helped schedule Sapphire's exercise time. This assistance was instrumental during the first 2 weeks of the program. However, during the latter 2 weeks of the program, Sapphire appeared internally motivated to complete her exercises. As stated by her mom, "The last week she [Sapphire] was dead set. No matter what, she would say we were going to make an hour or more each time."

Barriers and Usability Issues

During weeks 2 and 3, Sapphire's mom reported that there was a transition to a full-time school load delivered through the internet due to the COVID-19 quarantine. This required them to adjust their already strict schedule, which contributed to a lower amount of exercise. Sapphire reported that there was a learning curve to using the VR headset and games. It took 2 days of exercise before she and her mother figured out how to equip the headset comfortably over her glasses and adjust the headset appropriately to avoid pressure on the front of the face, which would cause uncomfortable temporary imprints on her forehead.

Although Sapphire reported that some games were perfectly suitable for playing in a wheelchair, other games—Dance Central and Racket Fury—required lowering the player height in either the game or Oculus Quest settings for optimal play. Default settings for the player height of these games caused many of the activities to be out of physical reach for Sapphire. Additionally, Dance Central instructed players to move their lower extremities, which Sapphire could not follow. Some games (eg, tennis) took up a considerable amount of floor space, and Sapphire and her mom had to coordinate play time in the living and dining rooms to avoid conflict with family activities (eg, television and meals).

When Sapphire would play a multiplayer game online, such as Rec Room, some other players would speak foul language. Sapphire was instructed by her mom to change game rooms until she was able to play the games with others who did not use offensive language. Similar to Johnny, Sapphire's mom stated that the heart rate monitor was sometimes not worn when playing, indicating that she spent more time exercising than what was recorded. The heart rate monitor had difficulties in syncing with the phone. Similar to Johnny, Sapphire also reported that the VR headset would have to be charged frequently due to the low battery life. She was using the headset for such long durations of play that the battery life (ie, 2.5 hours) could be expended within 1 or 2 days.

Discussion

Principal Results

This study was the first to objectively measure whether physically inactive youth with SB could engage in health-enhancing exercise behavior over a typical 1-month intervention period at home, using a state-of-the-art VR headset (ie, Oculus Quest). The study was implemented during a *shelter-in-place* order by the state due to the COVID-19 pandemic. Both participants reported that they engaged in no exercise behavior since COVID-19-related school closures, which was 2-3 weeks prior to starting the study. Yet, objective data indicated that both participants could use the headset to satisfy and exceed adult exercise guidelines of 150 minutes of moderate-intensity exercise across a 1-month period.

Although the levels of exercise obtained by participants were lower than 420 minutes of moderate-intensity exercise per week, which are part of the exercise guidelines for children [6], satisfying the adult exercise guidelines is an important finding for several reasons. First, adult exercise guidelines are widely prescribed as the minimum criteria for receiving health benefits from exercise. Second, the ability to maintain this level of exercise behavior will be important for living a healthy, active lifestyle as youth transition into adulthood. Third, benefits to health have been observed from even far lower levels of exercise behavior among youth with disabilities [7,11]. Fourth, levels of physical activity were higher than those reported in published exercise interventions for youth with SB [11] and other groups with physical disabilities [7]. There is a need to build on these findings and examine the effects of VR on critical health and function outcomes in this population.

The findings from the qualitative interviews helped explain the quantitative exercise data. First, participants reported that the amount of exercise they completed was heavily facilitated by the immersive and enjoyable movement that was involved with the VR games. Second, caregivers were identified as an agent that facilitated exercise behavior. Caregivers managed their child's exercise time and had a vital role in physically assisting them in setting up, adjusting, and using the VR system. Caregiver support appeared most influential for ensuring that participants equipped the heart rate monitor during exercise. Third, due to usability issues identified with the chest-worn heart rate monitor, the actual amount of exercise completed by each youth was likely far higher than what was objectively recorded by the VR system. In fact, both caregivers had to limit the participants from overusing the headset in order to not interfere with school-related work. Last, while participants did achieve an encouraging level of exercise behavior throughout the program, they identified minor usability issues that had to be resolved for an optimal gameplay experience.

The findings from this case study identified a promising protocol that can be used to engage youth with SB in health-enhancing levels of exercise behavior at home. A strength of this protocol was that youth were intrinsically motivated to play the VR games and, thus, did not need extensive behavioral coaching from research staff. Each participant was provided with one 20-minute coaching call at the start of the program and brief weekly reports. This required no more than a total of 1 hour and 40 minutes of time from research staff throughout the 1-month intervention. Thus, the protocol has minimal burden on research staff, making it highly useful for implementing in larger efficacy or effectiveness trials.

A second strength of this protocol was that participants were able to achieve a respectable volume of moderate-intensity exercise each week while playing games at a *normal* or *hard* level of difficulty. This is an important finding since the provided games typically had much higher levels of difficulty (eg, an easy, normal, hard, expert, and expert + difficulty setting). Given that higher game difficulties require faster and more dynamic movements and a higher level of executive function, participants still had substantial room for growth using the same intervention games. It took the researcher (BL), who had a background in professional-level gaming, approximately 3 months of daily play to master the hardest difficulty level of the provided games. The longevity of this type of program for youth with SB needs to be evaluated.

Future Recommendations

Lessons can be learned from this case study to inform the design of future efficacy or effectiveness trials. A bullet-point list of recommendations for future trials is shown below (see Textbox 2).

Caregiver

The most notable lesson this study was the importance of caregiver support. Caregivers were not provided with direct instructions to assist the participants with the VR exercise program. Yet, caregiver support naturally emerged as a critical facilitator of exercise behavior, and this should be the standard moving forward in a larger trial. Caregivers have substantial on influence their child's adherence to home-based programs [44,45]. In this study, caregivers were highly motivated to support their child in performing the VR games, which will not always be the case in a larger trial or, perhaps, a time when families are not quarantined to the home due to COVID-19. Future studies should include behavioral coaching strategies that can accompany VR gaming for not only the child participant but also for the caregiver.

Convenient Methods of Monitoring

A second notable lesson for future VR exercise trials is the need for a more convenient method of monitoring exercise data. The chest-worn heart rate monitor was inconvenient for the two participants and was identified as a barrier to adhering to the monitoring protocol. Caregivers often had to support the children in this process. This ultimately resulted in a lower objectively recorded volume of exercise than was performed. The monitor for this study was chosen for its compatibility with the VR Health Exercise Tracker app and for the evidence supporting its accuracy and precision. To enhance the likelihood that youth will comply with the monitoring procedures, future studies should consider incorporating more convenient monitoring devices, such as those worn on the wrist or arm. However, the cost and accuracy of wrist-worn devices vary considerably (67%-92% agreement with the Polar H7 monitor) [46]. Researchers and health professionals will have to weigh convenience, accuracy, and cost when choosing a monitoring device. Regarding exercise data, participants seemed to highly value the resultant calorie data over data for heart rate after each exercise session. Thus, an ideal monitoring protocol might include a valid and reliable wrist-worn device that can sync with an easy-to-use mobile phone app or the headset itself to monitor heart rate, calories, and exercise time.

Intervention Instructions

A third important lesson learned from this study was that participants and caregivers should be provided with instructions on how to properly set up, equip, and adjust the headset. Every time the Quest is turned on, the device requires the user to set and draw the maximum boundaries for the play area as a measure of safety. If an individual moves outside of these bounds, they no longer see the game through their headset and instead view their surrounding area. Participants can exercise through the headset in either a stationary space (~5 feet in diameter) or *room-scale* space (maximum 25 × 25 feet). A strength of the Quest is that it includes an *inside-out* tracking system (ie, built-in motion sensors within the headset that track arm and leg use) to allow the participant to easily change play locations, whereas most other HMDs of similar visual quality require external wall-mounted cameras and, thus, one designated play location. Nevertheless, families should be informed of these space requirements to avoid conflicts with their daily activities or interruptions to the exercise session. Notably, participants will need specific instructions or accessories (eg, counterweights, cushions, or elastic headband support) for adjusting the headset to the face of a child. Prolonged exercise bouts with the headset sometimes caused temporary, uncomfortable facial imprints. Included within the calibration of the headset is the ability to set the player's height. People who use wheelchairs must be sure to set the player height to eye-level height while sitting, or change the height level if a specific game does not have built-in features to accommodate a nonstanding player (eg, Dance Central). Otherwise, games will be out of arm's reach for the user.

Game Characteristics

This study also provided valuable insight on game characteristics that may be useful for developers and future interventionists. An encouraging finding was that most of the provided games for the Quest could be played using default game settings with only the participants' arms. The chosen games included most of the nonviolent active games that were available on Quest at the time of the study. Fortunately, most games for the Quest are based on dynamic arm movements instead of leg movements, which is a generally accessible feature for wheelchair users. Only a couple of games encouraged heavy use of the lower limbs or specific adjustments to default game settings (ie, lowering the player height) to be played successfully (ie, Dance Central and Racket: Nx). These games might require specific instructions for adaptation or modifications for accessibility if they are to be included in another trial.

The participants had slightly different preferences for game genres. Johnny spent most of his time playing single-player, action-oriented, and opponent-driven fighting games, whereas Sapphire generally preferred more cooperative and competitive multiplayer experiences. Providing participants with a variety of games within these genres might be important for promoting play. One notable exception was the rhythmic movement-to-music games, which were enjoyed by both participants, making this genre a potentially important one for development, modification, or inclusion for male and female youth with SB.

Although the youth were provided with both single-player and multiplayer games, the time spent playing multiplayer games appeared low, or participants did not choose to wear the monitoring system while playing these games. One participant perceived the online multiplayer community to be small. The other participant was exposed to inappropriate language by other users. Multiplayer games foster competition, cooperation, and friendships and, thus, have the potential to be more engaging over a long period of time. Further strategies for promoting and monitoring multiplayer gameplay are warranted.

Finally, the perceived realism of using the Quest appeared to be an important factor in participants' perceived enjoyment of, and likely their adherence to, the VR experience. This finding agrees with previous studies, whereby enjoyment was linked with VR immersion [47]. Participants reported that the experience "felt" real, not simply "looked" real, which implies that both the optical and proprioceptive systems were adequately immersed. This is likely due to the Quest having six degrees of freedom, a feature that is only included in the more sophisticated VR devices; a high-quality visual display; and no cable plug-in or external wires. Researchers and developers should aim to maintain or improve on this *realism* factor when developing future exercise-related VR trials or technology.

Textbox 2. Intervention recommendations for future trials.

- Caregivers should assist their child in using the hardware and software, when necessary, and encourage them to participate in the required amount of daily exercise.
- Exercise data, including caloric expenditure, should be objectively recorded by wrist- or arm-worn devices.
- Instructions should be provided to participants for setting up and adjusting the headset; video instruction could be especially helpful.
- Games that are usable for wheelchair users include the following: Beat Saber, AUDICA, Thrill of the Fight, Creed: Rise to Glory, Rec Room, Acron: Attack of the Squirrels!, Sports Scramble, and BOXVR.
- Strategies that encourage multiplayer gameplay should be developed and tested.

Limitations

Inherent within a case study design, the study findings should be interpreted with caution and will need to be confirmed in a larger efficacy trial. This study explored potential feasibility and usability issues and, thus, could not provide a more confirmatory result of acceptable or nonacceptable feasibility and usability, which could have been done through a single-subject research design. Similarly, we were unable to determine the impact or importance of the brief telecoach interactions with participants on exercise behavior. Another limitation was the lack of objective health and fitness laboratory measures, which would have provided estimates for treatment effects in a larger trial. Due to the COVID-19 quarantine, on-site testing was restricted at the university where the study was conducted. The COVID-19 quarantine may also have affected the level of engagement of both the participants and parents, as the study was implemented during a *shelter-in-place* order. In addition, the cost of the VR equipment may limit broad use for some populations.

Conclusions

This study tested the feasibility of using the latest state-of-the-art technology in consumer-available VR gaming among two youth with SB who used wheelchairs. Participants used the headset to safely achieve exercise guidelines, and this protocol holds promise for promoting physical activity on a larger scale among youth with SB. We identified several modifications to the VR exercise protocol and knowledge gaps that can be pursued in a future efficacy trial.

Acknowledgments

Funding for this project was provided by the Lakeshore Foundation in Homewood, Alabama, and the project was supported by the UAB/Lakeshore Foundation Research Collaborative.

Authors' Contributions

BL, DD, and MG developed the initial manuscript draft. All authors contributed equally to the later manuscript drafts. BH and BR assisted with participant recruitment. MJ constructed the figures and tables.

Conflicts of Interest

None declared.

References

1. Shin M, Besser LM, Siffel C, Kucik JE, Shaw GM, Lu C, Congenital Anomaly Multistate Prevalence and Survival Collaborative. Prevalence of spina bifida among children and adolescents in 10 regions in the United States. Pediatrics 2010 Aug;126(2):274-279. [doi: 10.1542/peds.2009-2084] [Medline: 20624803]
2. Dicianno BE, Karmarkar A, Houtrow A, Crytzer TM, Cushanick KM, McCoy A, et al. Factors associated with mobility outcomes in a national spina bifida patient registry. Am J Phys Med Rehabil 2015 Dec;94(12):1015-1025 [FREE Full text] [doi: 10.1097/PHM.0000000000000404] [Medline: 26488146]
3. Dosa NP, Foley JT, Eckrich M, Woodall-Ruff D, Liptak GS. Obesity across the lifespan among persons with spina bifida. Disabil Rehabil 2009;31(11):914-920. [doi: 10.1080/09638280802356476] [Medline: 19037774]
4. Oliveira A, Jácome C, Marques A. Physical fitness and exercise training on individuals with spina bifida: A systematic review. Res Dev Disabil 2014 May;35(5):1119-1136. [doi: 10.1016/j.ridd.2014.02.002] [Medline: 24612860]
5. Guidelines for the Care of People With Spina Bifida. 4th edition. Arlington, VA: Spina Bifida Association; 2018. URL: https://www.spinabifidaassociation.org/wp-content/uploads/Guidelines-for-the-Care-of-People-with-Spina-Bifida-2018.pdf [accessed 2020-08-27]

6. US Department of Health and Human Services. Physical Activity Guidelines for Americans. 2nd edition. Washington, DC: US Department of Health and Human Services; 2018. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2020-08-27]
7. Lai B, Lee E, Wagatsuma M, Frey G, Stanish H, Jung T, et al. Research trends and recommendations for physical activity interventions among children and youth with disabilities: A review of reviews. Adapt Phys Activ Q 2020 Apr 01;37(2):211-234. [doi: 10.1123/apaq.2019-0081] [Medline: 32131052]
8. Boudos RM, Mukherjee S. Barriers to community participation: Teens and young adults with spina bifida. J Pediatr Rehabil Med 2008;1(4):303-310. [Medline: 21791783]
9. Bellin MH, Zabel TA, Dicianno BE, Levey E, Garver K, Linroth R, et al. Correlates of depressive and anxiety symptoms in young adults with spina bifida. J Pediatr Psychol 2010 Aug;35(7):778-789 [FREE Full text] [doi: 10.1093/jpepsy/jsp094] [Medline: 19880443]
10. Mukherjee S. Transition to adulthood in spina bifida: Changing roles and expectations. ScientificWorldJournal 2007 Nov 26;7:1890-1895 [FREE Full text] [doi: 10.1100/tsw.2007.179] [Medline: 18060327]
11. Oliveira A, Jácome C, Marques A. Physical fitness and exercise training on individuals with spina bifida: A systematic review. Res Dev Disabil 2014 May;35(5):1119-1136. [doi: 10.1016/j.ridd.2014.02.002] [Medline: 24612860]
12. Jennett PA, Affleck Hall L, Hailey D, Ohinmaa A, Anderson C, Thomas R, et al. The socio-economic impact of telehealth: A systematic review. J Telemed Telecare 2003;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]
13. Bloemen MA, Verschuren O, van Mechelen C, Borst HE, de Leeuw AJ, van der Hoef M, et al. Personal and environmental factors to consider when aiming to improve participation in physical activity in children with spina bifida: A qualitative study. BMC Neurol 2015 Feb 10;15:11 [FREE Full text] [doi: 10.1186/s12883-015-0265-9] [Medline: 25886148]
14. Martin Ginis KA, Ma JK, Latimer-Cheung AE, Rimmer JH. A systematic review of review articles addressing factors related to physical activity participation among children and adults with physical disabilities. Health Psychol Rev 2016 Dec;10(4):478-494. [doi: 10.1080/17437199.2016.1198240] [Medline: 27265062]
15. Mat Rosly M, Mat Rosly H, Davis Oam GM, Husain R, Hasnan N. Exergaming for individuals with neurological disability: A systematic review. Disabil Rehabil 2017 Apr;39(8):727-735. [doi: 10.3109/09638288.2016.1161086] [Medline: 27108475]
16. Deutsch JE, Westcott McCoy S. Virtual reality and serious games in neurorehabilitation of children and adults: Prevention, plasticity, and participation. Pediatr Phys Ther 2017 Jul;29 Suppl 3:S23-S36 [FREE Full text] [doi: 10.1097/PEP.0000000000000387] [Medline: 28654475]
17. Sandlund M, McDonough S, Häger-Ross C. Interactive computer play in rehabilitation of children with sensorimotor disorders: A systematic review. Dev Med Child Neurol 2009 Mar;51(3):173-179 [FREE Full text] [doi: 10.1111/j.1469-8749.2008.03184.x] [Medline: 19191834]
18. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials. Neurosci Biobehav Rev 2017 Jul;78:34-43 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
19. Mitchell L, Ziviani J, Oftedal S, Boyd R. The effect of virtual reality interventions on physical activity in children and adolescents with early brain injuries including cerebral palsy. Dev Med Child Neurol 2012 Jul;54(7):667-671 [FREE Full text] [doi: 10.1111/j.1469-8749.2011.04199.x] [Medline: 22283557]
20. Rowland JL, Malone LA, Fidopiastis CM, Padalabalanarayanan S, Thirumalai M, Rimmer JH. Perspectives on active video gaming as a new frontier in accessible physical activity for youth with physical disabilities. Phys Ther 2016 Apr;96(4):521-532 [FREE Full text] [doi: 10.2522/ptj.20140258] [Medline: 26316530]
21. Meehan M, Insko B, Whitton M, Brooks FP. Physiological measures of presence in stressful virtual environments. ACM Trans Graph 2002 Jul;21(3):645-652. [doi: 10.1145/566654.566630]
22. Jones MB, Kennedy RS, Stanney KM. Toward systematic control of cybersickness. Presence 2004 Oct;13(5):589-600. [doi: 10.1162/1054746042545247]
23. McMahon DD, Barrio B, McMahon AK, Tutt K, Firestone J. Virtual reality exercise games for high school students with intellectual and developmental disabilities. J Spec Educ Technol 2019 Apr 11;35(2):87-96. [doi: 10.1177/0162643419836416]
24. Qian J, McDonough DJ, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: A systematic review. Int J Environ Res Public Health 2020 Jun 10;17(11):4133 [FREE Full text] [doi: 10.3390/ijerph17114133] [Medline: 32531906]
25. Rubin J, Chisnell D. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. 2nd edition. Indianapolis, IN: Wiley Publishing; 2008.
26. Fusch PI, Fusch GE, Ness LR. How to conduct a mini-ethnographic case study: A guide for novice researchers. Qual Rep 2017;22(3):923-941 [FREE Full text]
27. Hodge K, Sharp LA. Case studies. In: Smith B, Sparkes AC, editors. Routledge Handbook of Qualitative Research in Sport and Exercise. Abingdon, UK: Routledge; 2016:62-74.
28. Kelly D, Gibbons M. Marketing methodologies ethnography: The good, the bad and the ugly. J Med Market 2008 Aug 29;8(4):279-285. [doi: 10.1057/jmm.2008.18]

29. Creswell JW, Carroll Klassen A, Plano Clark VL, Clegg Smith K. Best Practices for Mixed Methods Research in the Health Sciences. Bethesda, MD: National Institutes of Health; 2011. URL: https://obssr.od.nih.gov/wp-content/uploads/2016/02/Best_Practices_for_Mixed_Methods_Research.pdf [accessed 2020-09-01]
30. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
31. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
32. Seaward BL, Sleamaker RH, McAuliffe T, Clapp JF. The precision and accuracy of a portable heart rate monitor. Biomed Instrum Technol 1990;24(1):37-41. [Medline: 2306564]
33. Speer KE, Semple S, Naumovski N, McKune AJ. Measuring heart rate variability using commercially available devices in healthy children: A validity and reliability study. Eur J Investig Health Psychol Educ 2020 Jan 10;10(1):390-404 [FREE Full text] [doi: 10.3390/ejihpe10010029]
34. Cicone ZS, Holmes CJ, Fedewa MV, MacDonald HV, Esco MR. Age-based prediction of maximal heart rate in children and adolescents: A systematic review and meta-analysis. Res Q Exerc Sport 2019 Sep;90(3):417-428. [doi: 10.1080/02701367.2019.1615605] [Medline: 31157608]
35. Virtual Reality Institute of Health and Exercise. URL: https://vrhealth.institute/ [accessed 2020-08-24]
36. Neuro-QoL: Measure development & research. HealthMeasures. URL: https://www.healthmeasures.net/explore-measurement-systems/neuro-qol/measure-development-research [accessed 2020-08-27]
37. Gershon RC, Lai JS, Bode R, Choi S, Moy C, Bleck T, et al. Neuro-QOL: Quality of life item banks for adults with neurological disorders: Item development and calibrations based upon clinical and general population testing. Qual Life Res 2012 Apr;21(3):475-486 [FREE Full text] [doi: 10.1007/s11136-011-9958-8] [Medline: 21874314]
38. Bertisch H, Rivara FP, Kisala PA, Wang J, Yeates KO, Durbin D, et al. Psychometric evaluation of the pediatric and parent-proxy Patient-Reported Outcomes Measurement Information System and the Neurology and Traumatic Brain Injury Quality of Life measurement item banks in pediatric traumatic brain injury. Qual Life Res 2017 Jul;26(7):1887-1899. [doi: 10.1007/s11136-017-1524-6] [Medline: 28271316]
39. Lai J, Nowinski C, Victorson D, Bode R, Podrabsky T, McKinney N, et al. Quality-of-life measures in children with neurological conditions: Pediatric Neuro-QoL. Neurorehabil Neural Repair 2012 Jan;26(1):36-47 [FREE Full text] [doi: 10.1177/1545968311412054] [Medline: 21788436]
40. Kozlowski AJ, Cella D, Nitsch KP, Heinemann AW. Evaluating individual change with the Quality of Life in Neurological Disorders (Neuro-QoL) short forms. Arch Phys Med Rehabil 2016 Apr;97(4):650-654.e8 [FREE Full text] [doi: 10.1016/j.apmr.2015.12.010] [Medline: 26740062]
41. Budgell B. Guidelines to the writing of case studies. J Can Chiropr Assoc 2008 Dec;52(4):199-204 [FREE Full text] [Medline: 19066690]
42. Fletcher AJ. Applying critical realism in qualitative research: Methodology meets method. Int J Soc Res Methodol 2016 Feb 29;20(2):181-194. [doi: 10.1080/13645579.2016.1144401]
43. Smith B, Sparkes AC. Introduction: An invitation to qualitative research. In: Smith B, Sparkes AC, editors. Routledge Handbook of Qualitative Research in Sport and Exercise. Abingdon, UK: Routledge; 2016:1-8.
44. Law M, King G. Parent compliance with therapeutic interventions for children with cerebral palsy. Dev Med Child Neurol 1993 Nov;35(11):983-990. [doi: 10.1111/j.1469-8749.1993.tb11580.x] [Medline: 8224565]
45. Schreiber JM, Effgen SK, Palisano RJ. Effectiveness of parental collaboration on compliance with a home program. Pediatr Phys Ther 1995;7(2):59-64. [doi: 10.1097/00001577-199500720-00003]
46. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
47. Shafer DM, Carbonara CP, Korpi MF. Factors affecting enjoyment of virtual reality games: A comparison involving consumer-grade virtual reality technology. Games Health J 2019 Feb;8(1):15-23. [doi: 10.1089/g4h.2017.0190] [Medline: 30199273]

Abbreviations

AFO: ankle-foot orthosis
AVG: active video gaming
COVID-19: coronavirus disease 2019
HMD: head-mounted display
Neuro-QoL: Quality of Life in Neurological Disorders
SB: spina bifida
UAB: University of Alabama at Birmingham
VO_2: oxygen volume

VR: virtual reality

Edited by G Eysenbach; submitted 26.05.20; peer-reviewed by J McCubbin, W Neumeier, L Grepo, T Cahill; comments to author 16.06.20; revised version received 20.07.20; accepted 11.08.20; published 03.09.20.

Please cite as:
Lai B, Davis D, Narasaki-Jara M, Hopson B, Powell D, Gowey M, Rocque BG, Rimmer JH
Feasibility of a Commercially Available Virtual Reality System to Achieve Exercise Guidelines in Youth With Spina Bifida: Mixed Methods Case Study
JMIR Serious Games 2020;8(3):e20667
URL: http://games.jmir.org/2020/3/e20667/
doi:10.2196/20667
PMID:32880577

Original Paper

Immersive Virtual Reality for the Reduction of State Anxiety in Clinical Interview Exams: Prospective Cohort Study

Brendan Joseph Concannon[1*], MSc; Shaniff Esmail[2*], PhD; Mary Roduta Roberts[2*], PhD

[1]Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[2]Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada

[*]all authors contributed equally

Corresponding Author:
Mary Roduta Roberts, PhD
Department of Occupational Therapy
University of Alberta
8205 - 114 St 2-64 Corbett Hall
Edmonton, AB, T6G 2G4
Canada
Phone: 1 780 492 2499
Fax: 1 780 492 4628
Email: mroberts@ualberta.ca

Abstract

Background: Immersive virtual reality (VR) with head-mounted display was used to determine if clinical interview simulation could reduce the anxiety levels of first-year occupational therapy (OT) students as they prepared for upcoming Objective Structured Clinical Examinations (OSCEs). Anxiety among health science students is a potential problem that may diminish their performance during OSCEs. This investigation aimed to fill the gap in the literature regarding the effectiveness of VR to reduce anxiety in OT students.

Objective: This investigation aimed to uncover the effectiveness of immersive VR in reducing state anxiety in OT students who were preparing for OSCEs.

Methods: A prospective, experimental, nonrandomized controlled trial compared levels of state anxiety, test anxiety, and academic self-efficacy in two groups of first-year OT students; these levels were measured at four different time points by self-reported psychometric scales, analyzed with a mixed factorial analysis of variance (ANOVA). Members of Phase 1 (NoVR) were not exposed to the VR simulation and acted as a control group for members of Phase 2 (YesVR), who were exposed to the VR simulation. VR simulation featured a virtual clinic and a standardized patient who students could interview in natural language. Measures of student study strategies and previous experience with VR were also recorded.

Results: A total of 49 participants—29 in the NoVR group and 20 in the YesVR group—showed that state anxiety had a rise-then-fall trend, peaking at the time point just before the OSCE. At that point, the YesVR students showed significantly less state anxiety than did the NoVR students ($t_{46.19}$=2.34, P=.02, Cohen d=0.65, ηp_2=0.105). The mean difference was 6.78 units (95% CI 0.96-12.61). In similar trends for both groups, student test anxiety remained relatively static across the time points, while academic self-efficacy continually increased. A moderate positive correlation was found for total time spent studying and peak state anxiety (NoVR r=.46, n=28, P=.01; YesVR r=.52, n=19, P=.02).

Conclusions: This investigation shows evidence of immersive VR's capability to reduce state anxiety in OT students preparing for clinical practical exams. Immersive VR simulation, used for the reduction of anxiety in health science students, can potentially lead to a future of positive mental health change from the virtual to the real world.

(JMIR Serious Games 2020;8(3):e18313) doi:10.2196/18313

KEYWORDS

virtual reality; VR; head-mounted display; HMD; immersive technology; occupational therapy; OSCE; simulation; psychology; anxiety

Introduction

Background

This investigation used immersive virtual reality (VR) to reduce anxiety among occupational therapy (OT) students who were preparing for a clinical practical exam. VR is a useful tool that can positively shape mental health. It utilizes a human-machine interface that immerses people into digitally rendered illusions that are multisensory in composition and projected by computer hardware. These illusions act as virtual environments, allowing people to condition themselves against symptoms of anxiety by undergoing VR exposure therapy (VRET), a form of systematic desensitization that facilitates mental fortification against a feared stimulus [1]. VRET allows for the training of affective regulation, while people are subjected to situational contexts that induce anxiety [1,2]. VRET can safely provide answers to inaccessible and intangible concepts by observing the responses of people who are subjected to fear- and anxiety-inducing stimuli, which would otherwise be considered too dangerous or unethical to perform in the real world [3]. Depending on the extent of a virtual system's designed capability, a person immersed within the virtual environment acts as a user who may encounter, interact, control, and modify the virtual world. The user's experience is *evoked* to improve their mental proficiency and habituate against fear and anxiety [1]. In this investigation, the anxiety under analysis is of the type that students may experience while preparing for clinical practical exams in health science programs.

Campus Anxiety: A Prevalent Problem

Anxiety is a feeling of uneasiness and worry, usually generalized and unfocused as an overreaction to a situation that is subjectively seen as menacing [4]. It is a theoretical construct, capable of being triggered in either general or specific situations, with *proneness* (ie, trait anxiety) representing the frequency and/or intensity of the response and *transitory* (ie, state anxiety) representing the momentary response at a specific point in time [5]. Trait anxiety is a stable construct that is associated with personality traits, influencing the degree to which a person's state anxiety response occurs within a specific point in time [5]. Spielberger and colleagues, in 1972 and 1978, developed a measure for test anxiety, which detects differences in test-specific personality traits between individuals.

Test anxiety is situation specific and associated with two components: (1) cognitive components that manifest symptoms of worry, due to student concerns regarding the outcome of an assessment, and task-irrelevant thinking, causing interference and shifting of attention to irrelevant content, and (2) affective components that manifest physiological reactions, such as increased heart rate and headache, nervousness, and tension (ie, emotionality) [6]. Self-centered worry cognitions and emotionality responses, which students may experience during testing situations, are potentially distracting and may disrupt concentration and attention, resulting in reduced performance on cognitive-intellectual tasks [7].

It is important to note that anxiety while under academic evaluation (ie, test anxiety) is normal, especially in situations where students have invested urgent and preparatory activities to win an ideal outcome. However, severe anxiety that causes students to "lock up," panic, or show an unexpected reduction in performance is a serious problem. Anxiety symptoms are expected to have a negative impact on student academic achievement, self-efficacy, and self-concept [8]. In a survey with 1099 responses from a Canadian university, 38.5% of the university students self-reported that they had suffered from test anxiety at some point during their studies, 20.5% of the surveyed students believed that professors were unable or unwilling to help, and 11.3% of the students indicated they would not seek help as this would act against social desirability [9]. Test anxiety on university campuses is associated with student burnout and increased rates of attrition [10].

Self-efficacy is the subjective belief in one's ability to successfully perform a given task [11]. Academic self-efficacy is of a specific type that pertains to academic situations, with greater levels being correlated with increased student class participation and exam performance at the higher grade-point average levels [12]. The relationship between academic self-efficacy and student anxiety, where the retention of academic self-efficacy is maintained by the suppression of state anxiety, was a primary outcome of interest in this investigation.

VR Versus Anxiety

VR is defined as a human-machine interface that allows users to *project* themselves into a computer-generated virtual environment, where specific objectives can be achieved [13]. A potential method for reducing anxiety involves the use of immersive VR, which allows people to learn how they would feel and respond—physiologically, tactfully, and procedurally—while interacting with virtual situations that the brain treats as real. Immersive VR can change a user's fear structure into an adapted one, removing the pathological kind that distorts reality and increases escapist tendencies [14]. The objective is to create an immersive virtual environment that simulates a specific testing situation, allowing users to learn how to adapt. This objective allows users the ability to develop anxiety tolerance by facilitating *inhibitory learning*, at both voluntary and involuntary levels, granting them resiliency after developing habituation from specific virtual situations to utilize in real-world situations [15]. *Inhibitory learning* is theorized to occur when anxiety suppression is achieved by neurobiological conditioning of the prefrontal motor cortex, amygdala, and hippocampus within the brain [16].

A fully immersive virtual environment allows users to accept and respond to artificial stimuli in a natural manner [13]. The component of VR that determines a user's perception of their surrounding virtual environment is their physical level of *immersion*, ranging from nonimmersive (eg, desktop computer showing the environment) to fully immersive (eg, head-mounted display VR) [17-19]. Interactivity—the degree to which a user's actions result in applicable responses within the virtual environment—is the second component of VR [19]. The third component is imagination: the degree to which a user feels he or she is within the virtual environment [19]. Presence is a subjective concept that defines the psychological degree to which a user understands where it is possible to act within the virtual environment [19]. People may feel deeply present in virtual environments when the experience makes them feel *involved* as they put their full attention on the virtual objectives [19].

These components influence VR's level of fidelity, which is the capability of a virtual environment to reflect the real world. High fidelity is achieved when a user's actions, senses, and thought processes in a virtual world closely or exactly resemble what would be transferrable to a similar situation in the real world.

In Gaggioli and colleagues' report on the use of VR to reduce workplace stress for teachers and nurses, VR was found to be more effective in treating anxiety than the traditionally accepted gold standard for psychological stress treatment, cognitive behavioral therapy [20]. Sports psychologists have developed immersive VR environments that train an athlete's mental concentration for sprinting events, depicting crowd-filled stadiums and competitors [21]. Designs of virtual hospital waiting rooms allow older adults the opportunity to be treated against anxiety-inducing stimuli, such as loud noises from distressed patients or crying infants [3]. This exposure could be employed to improve the efficacy of psychosocial therapy, such as cognitive behavioral therapy, with VR simulations resembling anxiety-inducing situations [3]. Kniffin and colleagues reported on diaphragmatic-breathing training for the retention of attentional control to enhance self-regulatory skills, which was tested on female students who were exposed to virtual avatars of aggressive males [2]. It was concluded that immersive VR was effective for the training of self-regulatory skills in this manner [2].

VR in Health Science

OT has recognized VR as a potential tool for treating clients, including those diagnosed with stroke, hemiparesis, musculoskeletal injury, brain injury, cerebral palsy, neurodevelopment disorders, geriatric limitation, mental health, and complex or chronic pain [22]. However, reports of VR's role in curricula for interprofessional skills training in students are typically peripheral in comparison [23]. OT will often employ Objective Structured Clinical Examinations (OSCEs), which are clinical practical exams that assess student core competencies, including procedural, clinical encounter, and history-taking skills [24]. OSCEs often feature standardized patients, who are actors trained to portray the characteristics of patients, giving students the opportunity to demonstrate their technical and nontechnical skills while in a controlled environment.

There are reports detailing the use of virtual standardized patients in medical education, allowing students to practice history-taking skills with reasonable differential diagnosis results [25]. However, there is a gap in the literature regarding the use of immersive VR systems for the reduction of anxiety in OT students. OT students are often expected to interview standardized patients during OSCEs while under formal evaluation, resulting in them potentially experiencing increased levels of state anxiety.

It is expected that an immersive VR simulation of a clinical practical exam will facilitate *inhibitory learning* in OT students, resulting in the suppression of their anxiety symptoms. These anxiety symptoms, which may impact student performance on clinical practical exams, are expected to be conditioned by immersive VR simulation, resulting in a reduction of student state anxiety levels and retention of academic self-efficacy levels. Immersive VR in this investigation is expected to demonstrate these positive changes in OT students and fill the gap in the literature regarding these conditions. The results of this investigation may inform future decisions of educational disciplines, considering the implementation of immersive VR for the reduction of performance anxiety, associated with clinical practical exams.

Aim of This Investigation

This investigation aimed to uncover immersive VR's effectiveness for reducing anxiety in OT students who were preparing for an OSCE. The human-machine interface utilized a head-mounted display to achieve an immersive experience, complete with speech-recognition software, allowing the use of natural language for conversing with a virtual standardized patient. This system was expected to optimize *inhibitory learning* for the facilitation of anxiety tolerance as detailed by Craske and colleagues' report [15]. Academic self-efficacy was also measured to determine its relationship with peak state anxiety.

Research Questions

This investigation was designed to answer the following research questions:

1. Does immersive VR simulation of a clinical practical exam (ie, OSCE) effectively reduce state anxiety in OT students when compared to a control group?
2. How is academic self-efficacy influenced by exposure to a VR simulation of an OSCE?

Expectations

This investigation was expected to reveal the following:

1. If the VR simulation is effective as a form of VRET, we predict a reduction in state anxiety scores at times when the VR simulation is available.
2. If the OSCE is an anxiety-inducing event, we predict a peak in state anxiety scores at the time closest to the OSCE. However, we also predict that students who are exposed to the VR simulation will show lower peak anxiety scores than controls.
3. We predict an inverse relationship between measures of state anxiety and academic self-efficacy.

Methods

Experimental Design

This investigation was a prospective, experimental, nonrandomized controlled trial, involving two groups of participants, each comprised of OT students in the first year of their program. Members of Phase 1 (NoVR) were not exposed to the VR simulation and acted as a control group for members of Phase 2 (YesVR), who were recruited in the following year and were exposed to the VR simulation. The OT program itself was consistent in terms of faculty practice opportunities. Facility infrastructure, teaching of the OSCE content, sequencing of the courses, scheduling of mandatory practice sessions, and the professors themselves remained the same between groups.

Unlike a blinded randomized design, this investigation allowed each group of participants to be aware of their status in the experimental process. Due to the critical opportunities for when the OSCEs became available, it was not possible for this investigation to feature a block-controlled trial. Had a standard randomized controlled trial been utilized, there would have been difficulties with randomizing the students to their designated conditions; withholding an intervention that has the potential to positively impact student performance and well-being is unethical. This investigation used a wait-list control design that allowed participants from the NoVR group to access the VR simulation at a future schedule, separate from the YesVR group.

By maintaining intact cohorts as separate control and intervention groups between the years, this acted as a strategy to reduce treatment diffusion, which may have occurred if both groups had been analyzed at the same time. A potential confounding factor between the groups involves the time spent on opportunities to practice for the OSCE, regardless of modality. It was considered that the YesVR group's state anxiety scores could have been influenced by additional time practicing with the VR simulation, regardless of its effectiveness. Therefore, it was important for both groups to log their total time spent in preparation for the OSCE to note any potential differences in time spent between groups.

Recruitment

Announcements providing details of the investigation were made by an announcer who was neutral to the investigation's outcome. The announcer was the same for each phase and was not a professor within the faculty; this was to minimize the compulsory pressure on students to participate. Announcements were made during a lecture to an OT class of 120 students for each phase. All OT students for each phase, who were in the first year of their program, were invited and eligible to participate. While students in both phases were informed of the availability of a survey package that became available for them to obtain and complete, Phase 2's announcement included additional information, explaining the risks associated with the use of immersive VR hardware.

Ethics

This investigation was approved by the Research Ethics Office of Research and Innovation, University of Alberta, Canada. After inspection, this investigation was deemed ineligible to record participant age and sex variables. This was to ensure participant anxiety scores would not be traceable by professors of the faculty, especially if that data belonged to participants who were unique to the student population and these participants could risk being identified. Census data pertaining to the demographics of the student body were allowable and have been provided in the results.

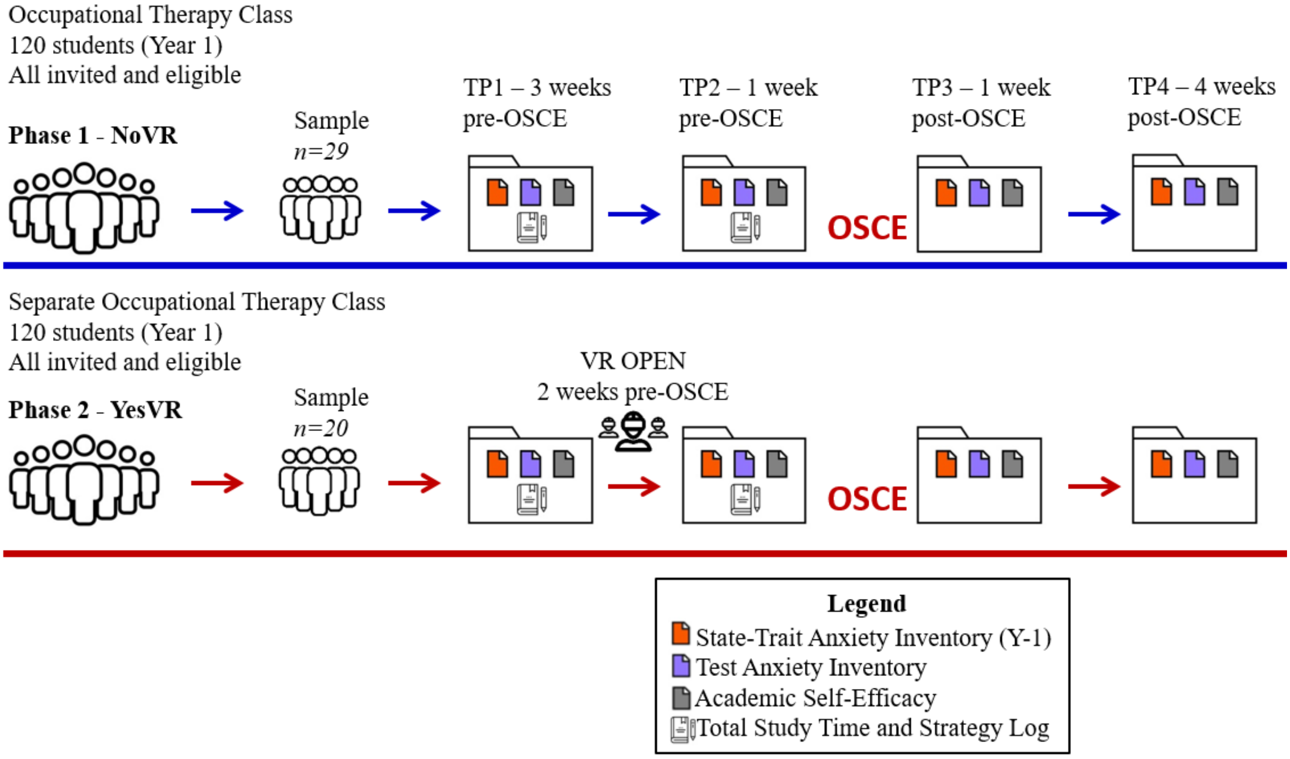
Experimental Process

Students were requested to obtain and complete a survey package that contained four separate sections, each to be completed at different time points (TPs): TP1, TP2, TP3, and TP4. Each section contained questionnaires that recorded primary and secondary outcome measures of this investigation. Once each section was complete, the participants were instructed to drop off each section at a secure mailbox as indicated within the package information guide. Note that Phase 2's (YesVR) package contained additional information about sign-up timeslots for immersive VR sessions, which would become available 2 weeks prior to their OSCE date. Tutorials on how to operate the VR hardware were provided by assistants, who were neutral to the investigation outcome and remained on standby at each appointed sign-up session. Each package section had been labelled with a specific completion date as follows:

1. TP1: 3 weeks before the OSCE.
2. VR sign-up became available for Phase 2 (YesVR) students only: 2 weeks before the OSCE.
3. TP2: 1 week before the OSCE.
4. TP3: 1 week after the OSCE.
5. TP4: 1 month after the OSCE.

Refer to Figure 1 for a summary of this investigation's experimental process.

Figure 1. Experimental process of this investigation. NoVR: subjects not exposed to the virtual reality simulation; OSCE: Objective Structured Clinical Examination; TP: time point; VR: virtual reality; YesVR: subjects exposed to the virtual reality simulation.



Phases 1 and 2: Primary Outcome Measures

State-Trait Anxiety Inventory Forms

The State-Trait Anxiety Inventory (STAI) consists of two scales, each comprised of 20 items that measure anxiety in adults; they are scored as a value ranging from 20 to 80, with higher scores being associated with stronger symptoms of anxiety [5]. This investigation utilized STAI Form Y-1—the State Anxiety (S-Anxiety) scale—which measures a participant's level of anxiety at a specific moment in time. The S-Anxiety scale has been found to be a "sensitive indicator of changes in transitory anxiety" as experienced by students exposed to stressors, such as job interviews or important school tests [5]. The STAI S-Anxiety scale was developed for use with college students and has shown a reliability stability of r<.62. Although reliability coefficients for the STAI have shown low-to-moderate scores, these stability coefficients are assumed for a state anxiety scale of this type, due to its expected ability to reflect differences in participant anxiety levels that are unique between each retesting situation [5]. Spielberger and colleagues [5] reported that normative Cronbach α coefficients for college students were .91 and .93 for males and females, respectively. For validity, the STAI S-Anxiety scale has been compared to other existing measures of state and trait anxiety in addition to contrasted groups, personality and adjustment tests, correlations with measures of academic aptitude, achievement, and investigations of the effects of different amounts and types of stress on S-Anxiety scores [5]. For college students, the Institute of Personality and Ability Testing (IPAT) Anxiety scale was compared to the STAI and showed validity correlation coefficients of r=.75 and r=.76 for females and males, respectively; however, a comparison between the STAI and the Taylor Manifest Anxiety Scale (TMAS) showed validity correlation coefficients of r=.80 and r=.79 for females and males, respectively [5]. The STAI has shown consistency in measuring essential qualities of anxiety, including apprehension, tension, nervousness, and worry [5]. The STAI Form Y-2—the Trait Anxiety (T-Anxiety) scale—measures a participant's general and long-standing level of anxiety, which was not featured in this investigation. Overall, the STAI has shown to be both a reliable and valid instrument for measuring state anxiety levels in college students.

Test Anxiety Inventory

The Test Anxiety Inventory, also known as the Test Attitude Inventory (TAI), is a self-reporting psychometric scale that measures individual differences in test anxiety as a situation-specific personality trait. It is comprised of 20 items that measure anxiety attributable to test situations and scored as a value ranging from 20 to 80, with higher scores being associated with stronger symptoms. TAI subscales include worry and emotionality as major qualities of test anxiety [26]. Although most normative data for TAI usage is based on general-purpose or multiple-choice tests, it allows for modification about specific tests or time periods accordingly [26]. The TAI is also useful as a measure of outcome for studies featuring test anxiety treatment [27-29]. The TAI scale was developed for use with college and graduate students and has shown a reliability stability of r=.80 for time periods varying between 2 weeks and 6 months [26]. Spielberger and colleagues [26] reported that the TAI Total scale showed uniformly high scores for both males and females (.92 or higher), with median α values for worry and emotionality subscales of .88 and .90, respectively. For validity, the TAI scale has been compared to other existing measures of test anxiety, including the Test

Anxiety Scale (TAS) and the Worry Emotionality Questionnaire (WEQ) [26]. For college students, the TAS and TAI comparison showed validity correlation coefficients of r=.83 and r=.82 for females and males, respectively; however, the TAI and WEQ-Emotionality comparison showed validity correlation coefficients of r=.85 and r=.77 for females and males, respectively [26]. Although there have been moderate positive correlations found when comparing the TAI with the STAI (r=.67 in males and r=.34 in females), the TAI was concluded not to measure or be comparable to state anxiety [26]. Overall, the TAI has been shown to be both a reliable and valid instrument for measuring test anxiety levels in college students.

Academic Self-Efficacy

In this investigation, academic self-efficacy was measured with the German Academic Self-Efficacy Scale (ASE), developed by Jerusalem and Satow in 1999 [30] as part of an extensive test battery to implement self-efficacy theory in schools of various grade levels up to and including trade school. Their instrument was developed by a combination of empirically proven concepts as outlined by Albert Bandura's Self-Efficacy Theory [11,31] and Jerusalem, Mittage, and Satow's research [30]. Their academic self-efficacy instrument is comprised of 7 items and showed internal consistency (ie, Cronbach α) of .73 when compared with the other tests that measured theoretically related constructs, such as optimism, helplessness, and social requirement expectations [32]. These theoretically related constructs were and were not related to academic self-efficacy (r values ranged from .27 to .51), resulting in theoretical correlations speaking for the criterion-oriented validity of the scale [30,32].

Phases 1 and 2: Secondary Outcome Measures

Each survey package contained a log template providing instructions on how to note study activities and durations in preparation for the OSCE. Participants were encouraged by the instructions to log each study activity and its duration on an ongoing basis. Participants were requested to provide only the times and activities that were outside their normal class and lecture sessions. Phase 2 (YesVR) participants were also requested to include their VR simulation session in their log and, if applicable, provide special notes as to why their VR session was incomplete had it ended prematurely. In addition, detailed instructions to sign up for optional interviews and focus groups were provided in the packages. Interviews and focus groups took place both before and after the OSCE, with the goal of determining student viewpoints on requirements for simulation effectiveness, immersiveness, feedback, and improvement, as well as mental mindset before and after the OSCE.

Phase 2 (YesVR) Only: Secondary Outcome Measures

A brief 5-item survey was an additional document available in the Phase 2 (YesVR) survey package, which allowed participants to define the amount of familiarity and ownership, if applicable, of immersive VR hardware they had experienced prior to the simulation as featured in this investigation. This survey established participant opinions regarding the following characteristics of VR environments: (1) VR features that they perceived to be the most important for establishing feelings of realism, (2) their preferred activities while using immersive VR, and (3) their prediction of immersive VR's potential as an educational tool for the future of education. It was important for the survey to specify the type of VR in each question and provide examples of VR headsets—Oculus Rift (Facebook Technologies), Vive (HTC Corporation), PlayStation VR (Sony Interactive Entertainment), Gear VR (Samsung Electronics Co), or Google Cardboard—so that any potential discrepancy between the interpretation of immersive and nonimmersive VR types was minimized. A copy of this survey is available in Multimedia Appendix 1. Overall, this survey was used to establish a baseline understanding of participants' attitudes toward immersive VR, prior to their involvement in the simulation as featured in this investigation.

Simulation Design

Overview

The simulation in this investigation included the following components:

1. A virtual environment depicting a health sciences clinic, rendered with Unity game engine software (Unity Technologies).
2. Two virtual avatars: the first was a virtual standardized patient who was located within the virtual environment and would respond to a user's questions; the second was a virtual exam evaluator who observed the user and would write notes into a clipboard during the interview process.
3. Speech-recognition software provided by IBM Watson, a question-answering engine linked with the virtual standardized patient.
4. VR (HTC Vive) and computer hardware that ran the software, allowing users to operate within the virtual environment itself.

Health Sciences Clinic

This project was designed and developed by an interdisciplinary team with experience in the use of VR learning objects for educational measurement, clinical evaluation, curriculum development, and assessment of student stress and anxiety. The project team members' expertise and their associated departments included the following: computing science, physical therapy, communication and science disorders, rehabilitation medicine, and OT. Throughout simulation development, multiple demos were performed to allow revisions, based on user feedback from each session.

Experts from the discipline of computer science were given a tour of the real-world health sciences clinic, allowing them to develop a virtual environment that closely resembled the OSCE setting as accurately as possible. The virtual environment was rendered with Unity game engine software and had two rooms: a hallway and an examination room (ie, doctor's office) that were separated by a door. The setting allowed users to move through the hallway, open the door, and walk into the doctor's office to meet the virtual standardized patient. The doctor's office included a patient examination table and a computer desk that was outfitted with a desktop computer and a miniature clock. At this point, a buzzer was sounded to signal the start of the OSCE and the miniature clock began to count down from 8 minutes. The avatar representing the exam evaluator was standing discretely in the corner of the room, writing notes into a clipboard throughout the interview process. The avatar representing the standardized patient was sitting in a chair, next to the patient examination table. Both avatars were programmed to maintain eye contact with the user. Refer to Figure 2 for a sample screenshot of the virtual health sciences clinic.

Figure 2. Screenshot of the virtual health sciences clinic.

OSCEs are described as follows, according to the Medical Council of Canada [33]: OSCEs are typically station oriented, attempting to resemble clinical scenarios with as much realism as possible. They are controlled and often feature trained actors who portray specific clinical patients in health-related situations. Although OSCE stations assess a variety of clinical competencies in students, assessments often focus on a student's ability to communicate with the patient, typically in an interview process with a history-taking approach. OSCE stations are timed and formally observed by evaluators who assess the student's performance.

Virtual Standardized Patient

The avatar representing the standardized patient was modelled to act as one of three different patients: Alex, Sam, or Jordan. They each had a different cause for their physical injury. The user could select a specific virtual patient or have one assigned randomly. They were voiced by the same voice actress (ie, standardized patient experience) and could respond to user questions or commands that were recognized and processed by IBM Watson's voice-recognition software. The avatar would raise her arms above her head when asked to do so, having a noticeable reduction in her range of motion for whichever limb was injured. She would respond in a respective manner to other physical actions, such as when she was asked to reach behind her back or touch her head.

Speech-Recognition Software

IBM Watson was linked to Unity, via an application programming interface, with a script that contained programming code to access the microphone located on the VR headset. The script then streamed audio data to the Watson speech-to-text service, allowing the virtual standardized patient to convert the verbal question of a user to text and check it with a list of applicable responses. If a user's verbal question matched an applicable response, the avatar would respond with an answer as previously voice-acted during her development. Her responses would vary depending on if she was Alex, Sam, or Jordan. Overall, the avatar was programmed to respond to an array of hard-coded questions that were divergent across the six components of health, including the physical, social, environmental, emotional, spiritual, and intellectual domains. She would also respond to other medical history questions, such as the reasoning of her doctor's referral or whether she was prescribed medication. She would respond appropriately when greeted. She would state that she did not understand a question when a user issued a verbal command that did not match any line of text from the list of applicable responses. Note that she was not programmed to understand or respond to convergent questions, such as "Can you tell me more?"

Participants were informed to reword their question or change the topic entirely if the virtual patient repeatedly failed to understand a question. A list of questions that Alex, Sam, or Jordan could understand and respond to is available in Multimedia Appendix 2.

VR and Computer Hardware

This investigation's VR hardware consisted of the HTC Vive, a consumer headset model with a built-in microphone, which allowed participants to interact within the virtual, health sciences clinic and converse with virtual standardized patients. The headset was supplemented with noise-cancellation headphones to reduce any real-world noise that could potentially contaminate the virtual clinic experience. The computer hardware was built using an Intel Core i5-6500K, 3.20 GHz CPU (central processing unit), NVIDIA GeForce GTX 1080 8 GB GPU (graphics processing unit), and 16 GB RAM.

Statistical Analysis

A 2 × 4 mixed factorial analysis of variance (ANOVA) was used to evaluate differences between and within the scores of each phase's STAI, TAI, and ASE. The two independent variables (ie, factors) were each designated with the following levels: 2 levels for representing each phase (ie, NoVR and YesVR) and 4 levels for representing each time point (ie, TP1, TP2, TP3, and TP4). Repeated-measures variables were corrected with Bonferroni *t* tests. Statistical significance was evaluated at α=.05, and a two-sided *P* value of .05 or less was considered to be statistically significant. Partial η^2 (ηp^2) effect size was checked to determine the ratio of variance accounted for by each effect and that effect plus its associated error variance within this ANOVA investigation. A ηp^2 effect was considered meaningful if found to be 0.06 or greater, indicating the effect explained 6% of the variance in the dependent variable. Protected *t* tests were used to compare each specific time point between phases as well as the total time spent preparing for the OSCE between phases. The Cohen *d* effect size was checked for protected *t* tests between phases. To account for conceivable events where immersive VR may have shown results that were opposite in direction to the expected results, such as state anxiety being increased in students due to the VR intervention itself, analysis checks for differences were two-tailed. Pearson correlation coefficients were performed between both phases' peak anxiety time points and study times, plus total peak anxiety and total academic self-efficacy. The peak anxiety time point was expected to be TP2, as it had the closest temporal distance to the OSCE of 1 week.

Power Analysis

A power analysis was calculated using G*Power (Heinrich-Heine-Universität) [34] to determine the total sample size needed for each ANOVA—2 groups, 4 measurements, repeated measures, and between factors—and its associated scale (ie, STAI, TAI, and ASE). A power analysis of 0.8 with α=.05, expecting a large effect size for an ANOVA (*f*=0.40), required a total sample size of 34, with 17 participants required per group. Each ANOVA score result was considered meaningful if both the sample size and effect size thresholds were met (ie, 17 participants per group and *f*≥0.40, respectively).

Results

Overview

A total of 49 OT students participated in the study: 29 for Phase 1 (NoVR) and 20 for Phase 2 (YesVR). Although the response rate was 100% for package submissions, Phase 2 (YesVR) had 1 participant out of 20 (5%) fail to submit a completed TAI survey. Only 1 participant out of 20 (5%) from Phase 2 (YesVR) reported to have suffered from simulation sickness, yet was still able to complete the simulation. The majority of students in Phase 2 utilized the VR simulation for a single 15-minute session, with some having multiple sessions, which resulted in a mean VR simulation time of 17.32 minutes (SD 7.52) per student. Although this investigation was unable, ethically, to obtain participant demographics, census data from the OT student body are available in Table 1. The main statistical analysis results are provided in Table 2.

State Anxiety

Figures 3 and 4 shows student state anxiety across time. The results of the 2 × 4 mixed ANOVA showed there was no significant main effect for phase ($F_{1,47}$=0.276, *P*=.60, ηp^2=0.006) on state anxiety scores, with NoVR (mean 40.78, SD 12.82) and YesVR (mean 39.54, SD 10.04) performing similarly overall. However, there was a significant difference in state anxiety scores between phases at TP2, with NoVR showing greater anxiety scores (mean 48.03, SD 12.67) than YesVR (mean 41.25, SD 7.54) ($t_{46.19}$=2.34, *P*=.02, Cohen *d*=0.65, ηp^2=0.105). The mean difference was 6.78 units (95% CI 0.96-12.61). Cronbach α values for the participant samples at TP2 were .95 and .87 for NoVR and YesVR groups, respectively.

Table 1. Student census data.

Year	Total class size (students), N	Aged 18-23 years, n (%)	Aged 24-29 years, n (%)	Aged 30-34 years, n (%)	Aged 35-39 years, n (%)	Identified as female, n (%)	Identified as male, n (%)
2017	123	13 (10.5)	106 (86.2)	3 (2.4)	1 (0.8)	109 (88.6)	14 (11.4)
2018	121	20 (16.5)	88 (72.7)	12 (9.9)	1 (0.8)	110 (90.9)	11 (9.1)

Table 2. Statistical analysis results.

Variables	Phase 1 (NoVR[a]), mean (SD)	Phase 2 (YesVR[b]), mean (SD)	*df*	ANOVA[c]			*t* test	*P* value (Cohen *d*[d])
				F test	*P* value	Partial η^2		
STAI[e](Form Y-1) score at each time point (TP)								
TP1	44.34 (12.38)	40.70 (11.06)	47	—	—	—	1.06	.30
TP2	48.03 (12.67)	41.25 (7.54)	46.19[f]	—	—	—	2.34	.02 (0.65)
TP3	39.55 (10.60)	41.45 (11.69)	47	—	—	—	–0.59	.56
TP4	31.17 (9.17)	34.75 (8.76)	47	—	—	—	–1.37	.18
STAI (Phase)	—	—	47	0.28	.60	0.006	—	—
STAI (Time)	—	—	3	18.40	<.001	0.281	—	—
STAI (Intercept)	—	—	3	4.12	.008	0.081	—	—
TAI[g]at each TP								
TP1	41.59 (14.07)	41.89 (12.99)	46[h]	—	—	—	–0.77	.94
TP2	42.72 (13.55)	41.21 (11.58)	46[h]	—	—	—	0.40	.69
TP3	40.62 (13.62)	41.68 (11.48)	46[h]	—	—	—	–0.28	.78
TP4	40.52 (14.40)	41.79 (12.50)	46[h]	—	—	—	–0.32	.75
TAI (Phase)	—	—	46[h]	0.01	.94	<0.001	—	—
TAI (Time)	—	—	3	0.67	.57	0.014	—	—
TAI (Intercept)	—	—	3	1.57	.20	0.033	—	—
ASE[i]at each TP								
TP1	18.90 (2.43)	19.20 (3.12)	47	—	—	—	–0.38	.70
TP2	18.93 (2.37)	19.45 (3.10)	47	—	—	—	–0.66	.51
TP3	19.79 (2.64)	19.80 (2.93)	47	—	—	—	–0.01	.99
TP4	20.03 (2.76)	20.55 (2.72)	47	—	—	—	–0.65	.52
ASE (Phase)	—	—	47	0.22	.64	0.005	—	—
ASE (Time)	—	—	2.62[j]	8.98	<.001	0.160	—	—
ASE (Intercept)	—	—	2.62[j]	0.40	.73	0.008	—	—

[a]NoVR: subjects not exposed to the virtual reality simulation.

[b]YesVR: subjects exposed to the virtual reality simulation.

[c]ANOVA: analysis of variance.

[d]Cohen *d* effect size is only reported for time point 2 (TP2).

[e]STAI: State-Trait Anxiety Inventory.

[f]Levene’s Test for Equality of Variances found equal variances not assumed; thus, *df* was changed accordingly.

[g]TAI: Test Attitude Inventory.

[h]Phase 2 (YesVR) had 1 participant fail to complete TAI surveys.

[i]ASE: Academic Self-Efficacy Scale.

[j]Mauchly’s test found sphericity assumption violated; thus, *df* was corrected using Greenhouse-Geisser estimates of sphericity (ε=.88).

Figure 3. Student state anxiety at all time points (TPs); error bars represent standard error. The dotted line represents the time when the Objective Structured Clinical Examination (OSCE) took place. NoVR: subjects not exposed to the virtual reality simulation; YesVR: subjects exposed to the virtual reality simulation.

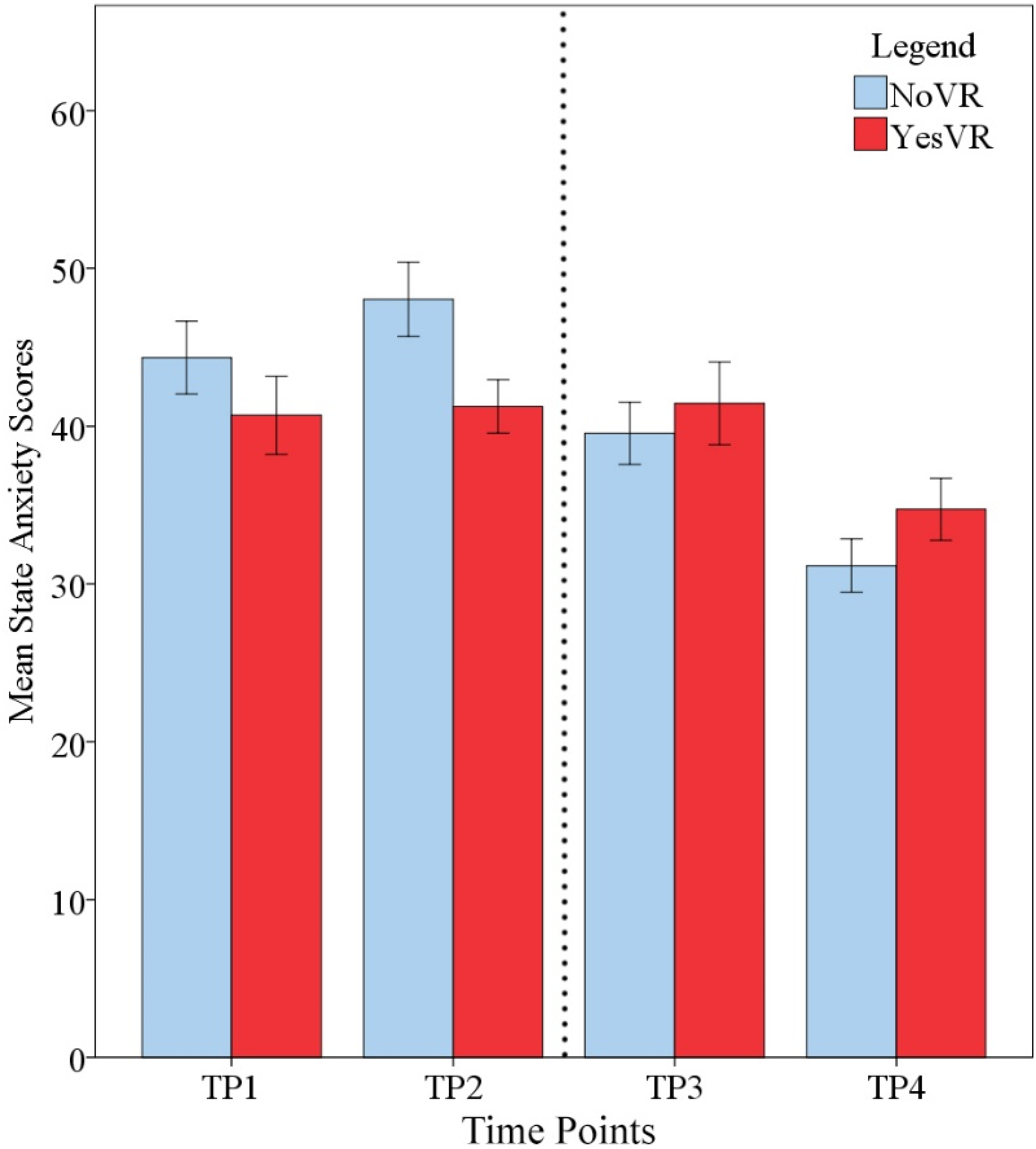
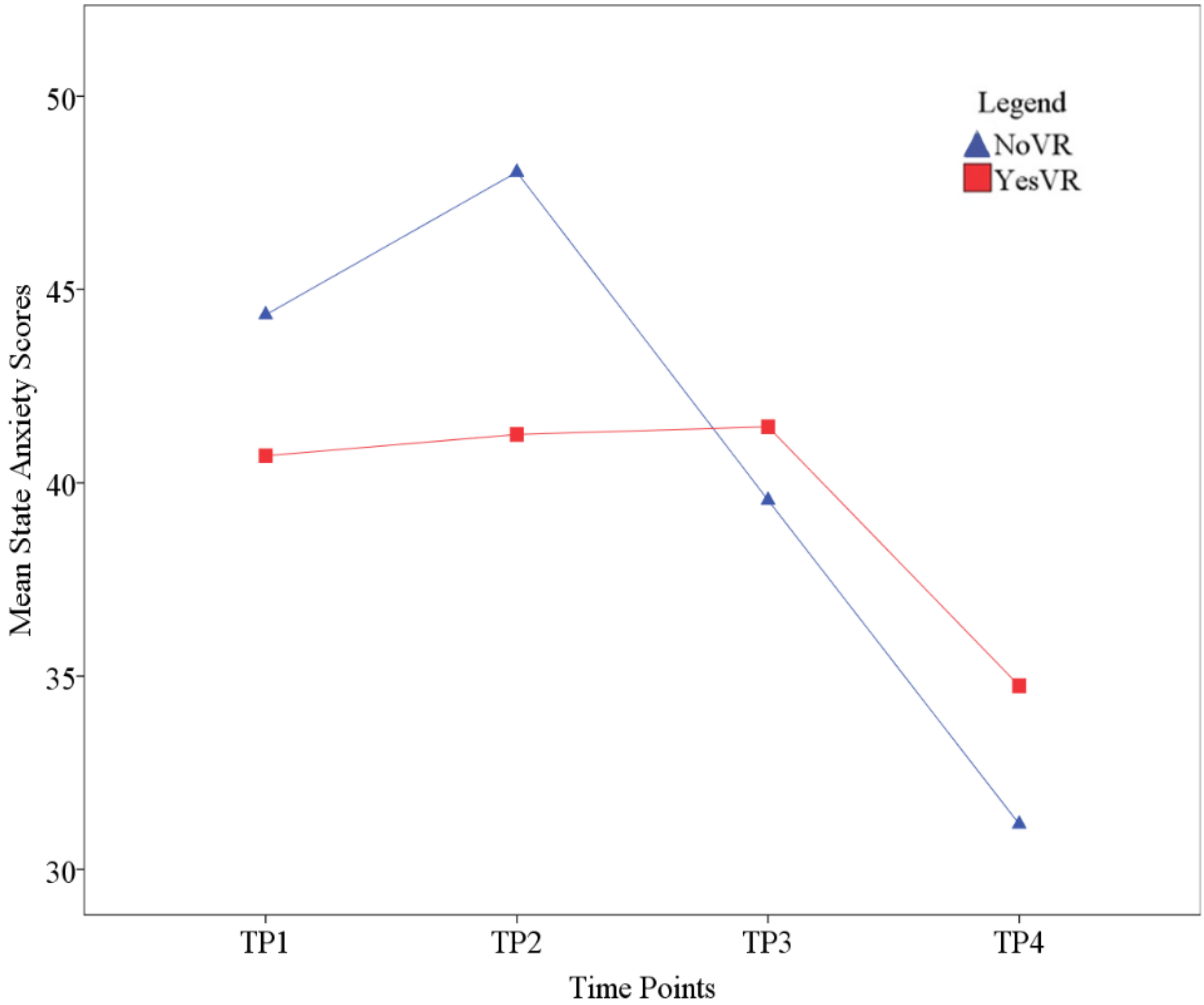


Figure 4. Student state anxiety at all time points (TPs). NoVR: subjects not exposed to the virtual reality simulation; YesVR: subjects exposed to the virtual reality simulation.



There was a significant effect for time on state anxiety scores ($F_{3,141}$=18.40, P<.001, ηp^2=0.281, f=0.6), with participants showing a rise-then-fall trend in mean state anxiety scores across the time points (TP1=42.86, TP2=45.27, TP3=40.33, and TP4=32.63). Pairwise comparisons that were corrected with Bonferroni *t* tests and CI adjustments showed a significant difference between TP1 (mean 42.86, SD 11.88) and TP4 (mean 32.63, SD 9.09) (P<.001), with a mean difference of 9.56 units (95% CI 5.12-14.00). There was also a significant difference between TP2 (mean 45.27, SD 11.29) and TP4 (P<.001), with a mean difference of 11.68 units (95% CI 7.30-16.06). Lastly, a significant difference was found between TP3 (mean 40.33, SD 10.98) and TP4 (P<.001), with a mean difference of 7.54 units (95% CI 2.94-12.14).

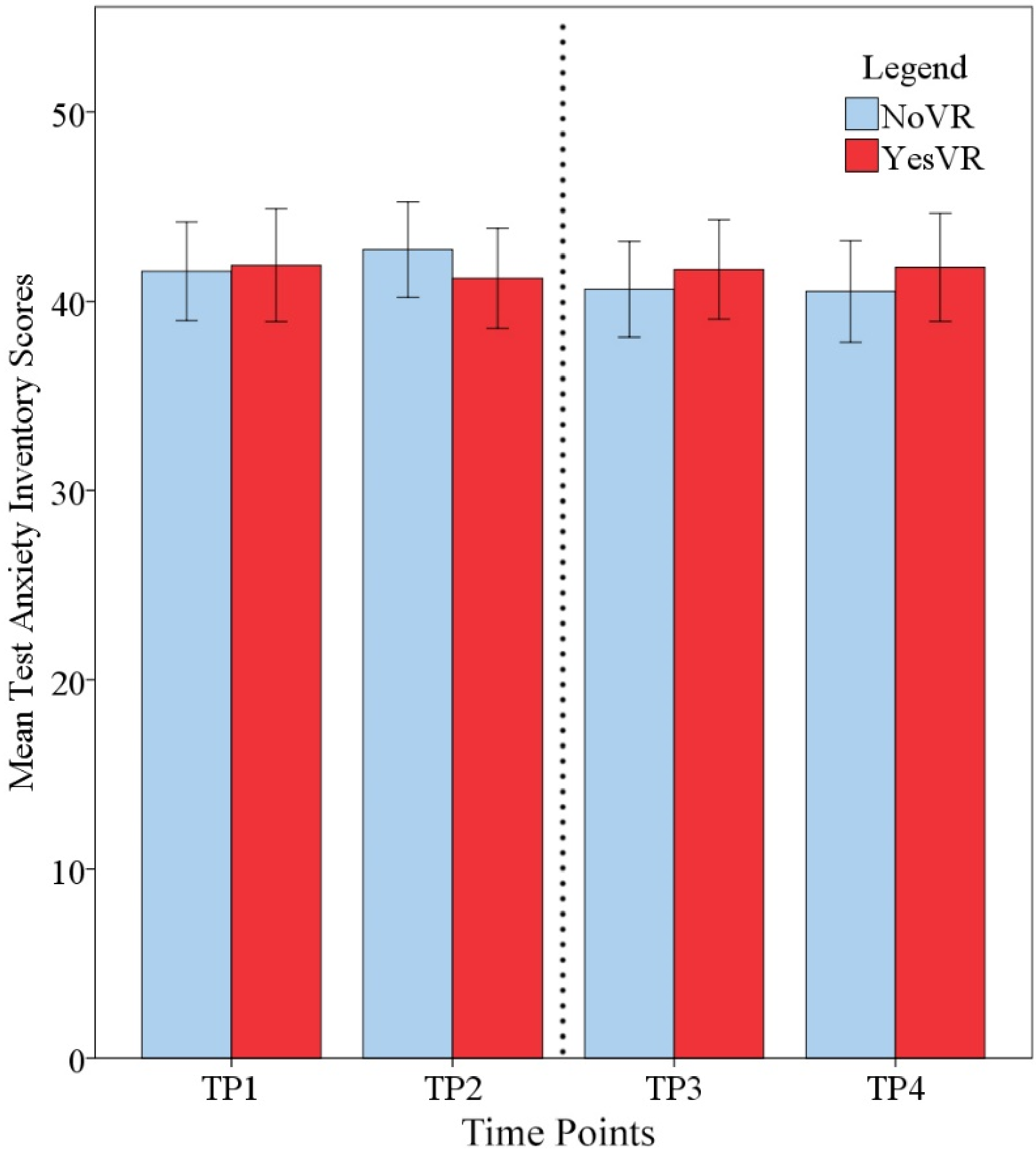
There was a significant interaction between time and phase in terms of state anxiety scores ($F_{3,141}$=4.12, P=.008, ηp^2=0.081, f=0.25). Descriptive statistics showed that NoVR participants showed greater state anxiety scores for TP1 (mean 44.34, SD 12.38) and TP2 (mean 48.03, SD 12.67) than did YesVR participants for TP1 (mean 40.70, SD 11.06) and TP2 (mean 41.25, SD 7.54). However, the opposite pattern occurred at TP3 and TP4 with NoVR participants, where they showed lower state anxiety scores (mean 39.55, SD 10.60, and mean 31.17, SD 9.17, respectively) than did YesVR participants (mean 41.45, SD 11.69, and mean 34.75, SD 8.76, respectively).

These results show that NoVR participants' state anxiety had a rise-then-fall trend, peaking at the time point just before the OSCE. At that point, students who had access to the VR clinical simulation showed less anxiety than did the control students. The YesVR participants' state anxiety showed no change in state anxiety scores from TP1 to TP3, before falling at TP4.

Test Anxiety

Figure 5 shows student test anxiety across time. The results of the 2 × 4 mixed ANOVA showed there was no significant main effect for phase ($F_{1,46}$=0.005, P=.94, ηp^2<0.001) on test anxiety inventory scores, with NoVR (mean 41.36, SD 13.76) and YesVR (mean 41.65, SD 11.74) participants performing similarly overall. There was no significant effect for time on test anxiety inventory scores ($F_{3,138}$=0.674, P=.57, ηp^2=0.014), with participants showing a similar level of scores across the time points (TP1=41.71, TP2=42.13, TP3=41.04, and TP4=41.02). There was no significant interaction between time and phase in terms of test anxiety inventory scores ($F_{3,138}$=1.57, P=.20, ηp^2=0.033).

Figure 5. Student test anxiety at all time points (TPs); error bars represent standard error. The dotted line represents the time when the Objective Structured Clinical Examination (OSCE) took place. NoVR: subjects not exposed to the virtual reality simulation; YesVR: subjects exposed to the virtual reality simulation.

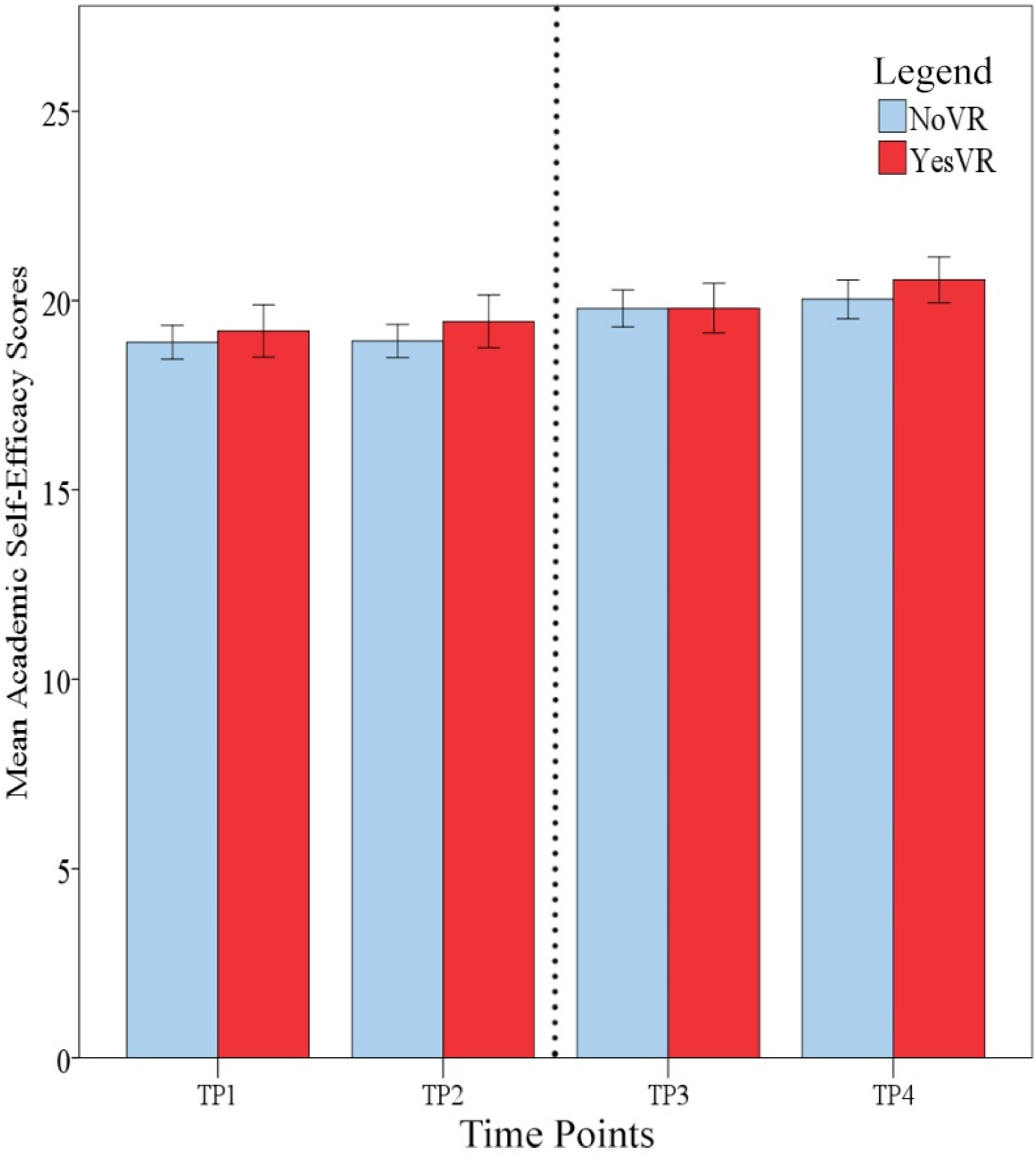


These results show that the students' test anxiety scores had remained relatively static throughout their participation in the OT program, whether they had access to VR or not.

Academic Self-Efficacy

Figure 6 shows student academic self-efficacy across time. The results of the 2 × 4 mixed ANOVA showed there was no significant main effect for phase ($F_{1,47}$=0.217, P=.64, ηp^2=0.005) on academic self-efficacy scores, with NoVR (mean 19.41, SD 2.57) and YesVR (mean 19.75, SD 2.96) participants performing similarly overall. Cronbach α values for the participant samples at TP2 were .61 and .74 for NoVR and YesVR groups, respectively.

Figure 6. Student academic self-efficacy at all time points (TPs); error bars represent standard error. The dotted line represents the time when the Objective Structured Clinical Examination (OSCE) took place. NoVR: subjects not exposed to the virtual reality simulation; YesVR: subjects exposed to the virtual reality simulation.



Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of time: χ^2_5=13.32 (*P*=.02). Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=.88). There was a significant effect for time on academic self-efficacy scores ($F_{2.62,123.32}$=8.98, *P*<.001, ηp^2=0.160, *f*=0.41), with participants showing an increase in mean academic self-efficacy scores across the time points (TP1=19.02, TP2=19.14, TP3=19.80, and TP4=20.24). Pairwise comparisons that were corrected with Bonferroni *t* tests and CI adjustments showed a significant difference between TP1 (mean 19.02, SD 2.70) and TP4 (mean 20.24, SD 2.73) (*P*=.001), with a mean difference of 1.24 units (95% CI 0.45-2.04). There was also a significant difference between TP2 (mean 19.14, SD 2.68) and TP4 (*P*=.002), with a mean difference of 1.10 units (95% CI 0.32-1.89).

There was no significant interaction between time and phase in terms of academic self-efficacy scores ($F_{2.62,123.32}$=0.40, *P*=.73, ηp^2=0.008). There was a significant moderate negative correlation between mean, total, academic self-efficacy (mean 19.14, SD 2.68) and mean, total, peak state anxiety scores at TP2 (mean 45.27, SD 11.30) (*r*=–.42, n=49, *P*=.003).

These results show that student academic self-efficacy had continually increased throughout their participation in the OT program, whether they had access to VR or not, yet it was inversely related to state anxiety at the peak anxiety time point.

State Anxiety, Study Time, and Strategies

Figures 7 and 8 show respective NoVR and YesVR phase correlations of student peak anxiety levels in relation to their total study times. For the NoVR participants, there was a significant moderate positive correlation between total study time and student state anxiety scores at TP2 (*r*=.46, n=28, *P*=.01). For the YesVR participants, there was a significant moderate positive correlation between total study time and student state anxiety scores at TP2 (*r*=.52, n=19, *P*=.02).

Figure 7. Correlation of student state anxiety and total study time at time point 2 (TP2) for subjects not exposed to the virtual reality simulation (NoVR). The line represents Pearson *r*=.46.

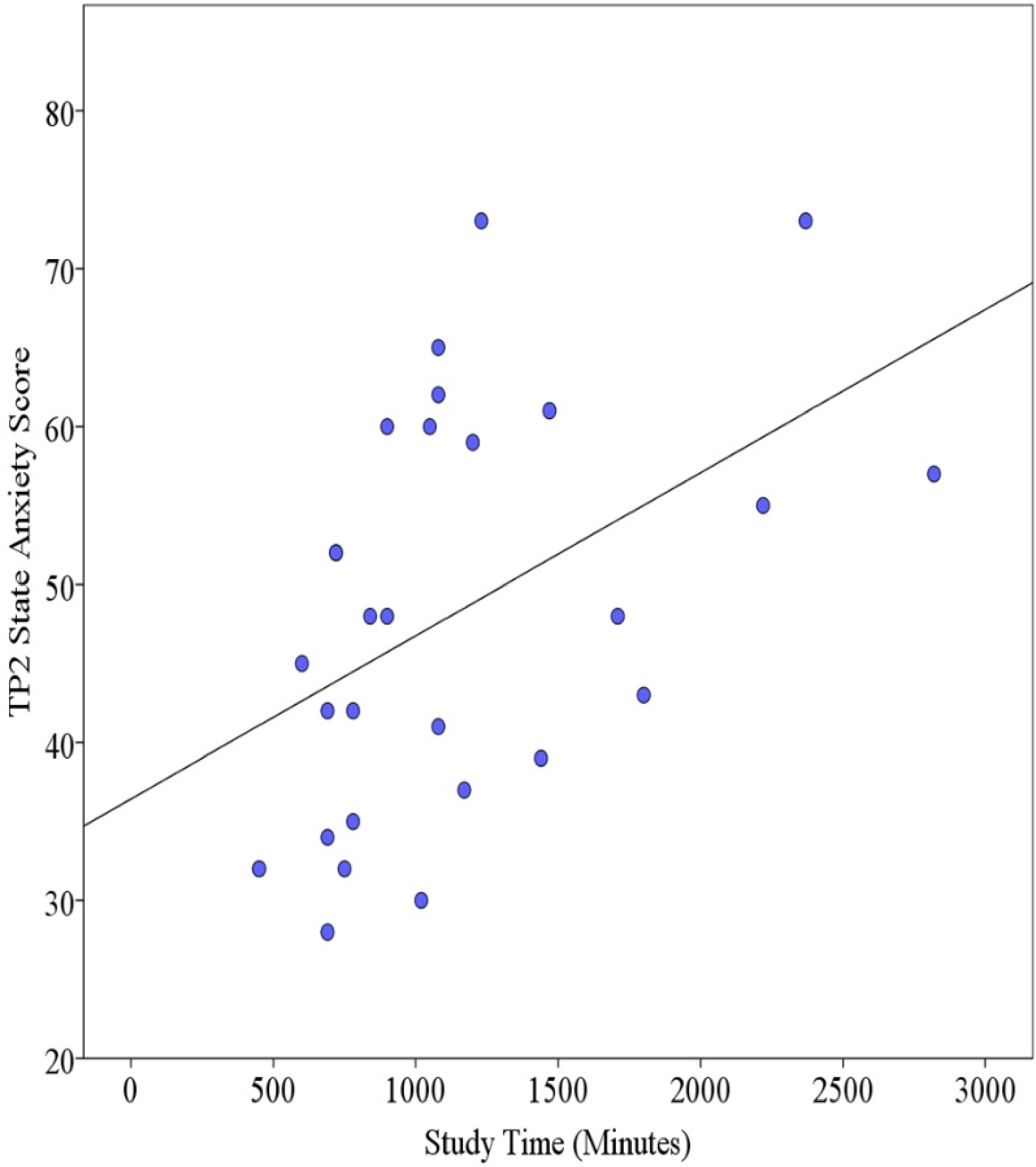
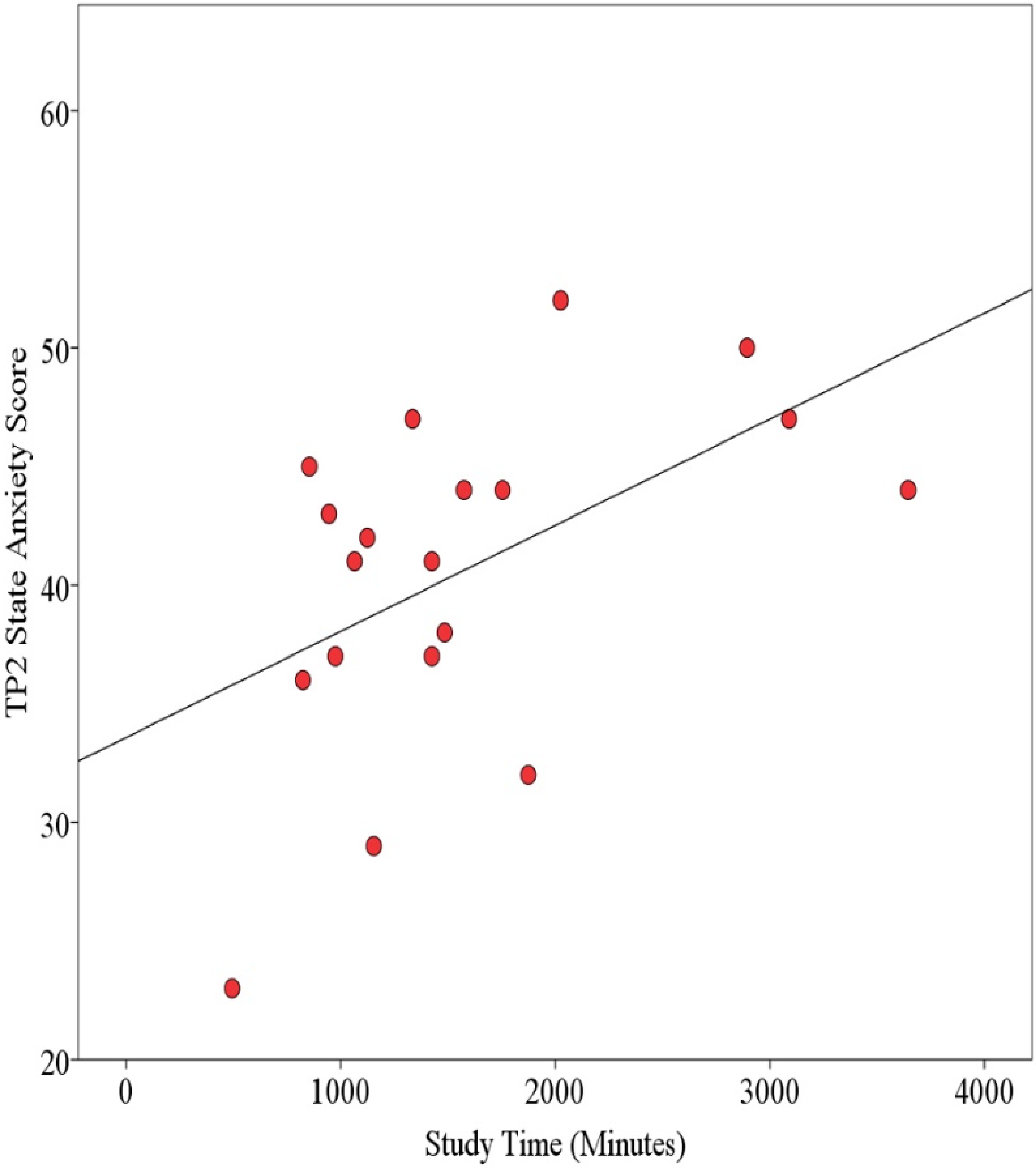


Figure 8. Correlation of student state anxiety and total study time at time point 2 (TP2) for subjects exposed to the virtual reality simulation (YesVR). The line represents Pearson *r*=.52.



There was no significant difference in total study time per week between the NoVR group (mean 383 minutes/week, SD 191) and the YesVR group (mean 315 minutes/week, SD 166) (t_{45}=1.27, *P*=.21, Cohen *d*=0.38). The top study strategies used by the NoVR group, as well as the mean percentages of time spent studying, were *hands-on group practice* (43.4%), *making and reviewing notes* (17.9%), *hands-on individual practice* (15.3%), and *reviewed e-class resources* (8.4%). The YesVR group used the same top study strategies in the same order: *hands-on group practice* (46.6%), *making and reviewing notes* (24.8%), *hands-on individual practice* (21.2%), and *reviewed e-class resources* (4.2%). The fifth-most common study strategy used by the NoVR group was *self-talk* (3.3%), while that of the YesVR group was *using the VR simulation* (1.1%).

The results show that excess time spent studying for clinical practical examinations was related to an increase in state anxiety. In summary, the top four study strategies used by the two groups were of the same type and were in the same order, yet the fifth-most common strategy for the NoVR group employed *self-talk* strategies, while the YesVR group used *VR simulation*.

Interview and Previous Experience with VR Survey Responses

In short summary, common student statements regarding realism and presence of the VR simulation from our interviews and focus groups included the following:

> *Feels like I'm really taking the exam.*
>
> *I almost freaked out when I saw the examiner in the room.*
>
> *The virtual reality let me experience the clinical exam in advance, which helped a lot.*
>
> *I'd love to work with more VR in the future.*

Refer to Table 3 for student responses to the Previous Experience with VR Survey.

Table 3. Previous Experience with VR Survey responses.

Item	Response (N=20), n (%)
No previous experience.	12 (60)
Neither owned nor knew someone who owned an immersive VR (virtual reality) system.	17 (85)
Main interest for immersive VR was training and education.	13 (65)
Quality of graphics is the most important requirement to make VR simulation feel the most realistic.	19 (95)
Both quality of graphics and ease of use and interaction with the virtual environment are the most important requirements to make VR simulation feel the most realistic.	17 (85)
Prefer to use immersive VR for professional work and education.	19 (95)
Use VR for both professional work and education and learning new skills.	14 (70)
Believed VR would have at least moderate potential as an educational tool within the next 10 years.	10 (50)
Believed VR would be "the way of the future" in the next 10 years.	7 (35)

Discussion

Overview

This investigation shows evidence of immersive VR's potential to reduce anxiety in OT students during their peak state anxiety time point as they prepared for an upcoming OSCE. However, this investigation did not fulfill the requirements of a rigorous randomized controlled trial. Thus, causation of immersive VR's effectiveness for the reduction of anxiety in OT students in training cannot be inferred. Favorable demonstrations of immersive VR's ability to reduce anxiety must consider that cognitive and affective responses, within the confines of a virtual world, could vary from those observed in the real world. The artificial nature of OSCEs are designated to act as simulations themselves, and despite their attempts to represent situations within the real world, they are not the real world [35]. Such can be stated for immersive VR simulation and its potential to optimize mental preparedness in medical students for the real world. Despite these remarks, previous research on immersive VR has demonstrated its capability for reducing state anxiety within the field of health sciences [20]. The results of this investigation reflected those trends.

Main Findings

State anxiety in the OT students was found to be different between phases only at the time point located 1 week prior to the OSCE (ie, TP2), which encompassed the VR intervention for the purpose of *inhibitory learning* to occur. It cannot be ruled out that differences in participant characteristics, covariates, and coursework, or even the qualities of the students themselves who agreed to participate, may have been responsible for the difference in state anxiety observed. It also cannot be stated that student performance anxiety levels were reduced during their actual performance in the OSCE event itself, since no measures were taken at that time. Gaggioli and colleagues' workplace stress report [20] showed a main effect for the reduction of anxiety in their VR experimental group, yet their study featured multiple VR sessions—eight treatment sessions—while the mode of this investigation featured only one. Although the state anxiety scores between groups was not significantly different at the first time point, the mean state anxiety being greater among the NoVR participants at that time may arguably have been the instigating difference for the second time point, despite test anxiety traits being similar for both groups.

First impressions of VR's influence in this investigation would appear to be mixed. State anxiety levels were not found to be significantly reduced for the YesVR group across all time points. At the first time point, the VR intervention had not been applied; thus, a difference in state anxiety levels at this point would have indicated a confounding difference between the groups. For the second time point, after the VR intervention had been applied, the YesVR group demonstrated an absence of an increase in state anxiety. This may have been influenced by other potential confounders. Although this investigation was able to consider test anxiety traits, faculty and coursework consistency, and study strategies, these potential confounders were not rigorously controlled and may have been the prime determinants of state anxiety differences. At the third time point, both groups' state anxiety levels returned to baseline levels, which were not significantly different from one another. This was after the OSCE had ended, yet the students were still working on other coursework, most likely subjecting them to different stressors than what was shown in the VR simulation. At the fourth time point, student anxiety levels were at their lowest, because this was when they were starting their next term with new coursework, when stressors have not yet manifested. These third and fourth time points were when the use of VR showed no retention of reduced state anxiety. These observations suggest that VR's ability to reduce state anxiety is considerable when it matters most; a specific intervention is given 2 weeks before the peak anxiety time point for a specific stressor. However, VR's ability to reduce state anxiety was not transferrable across different stressors, nor did it show a prolonged reduction in state anxiety over a longer period of time. The implication of these anxiety levels implies that OT departments, considering the incorporation of OSCEs for performance assessments, should expect students to experience increasing levels of state anxiety, especially when their OSCE appointments draw near. It is recommended that systems be implemented to mitigate this potential increase in state anxiety.

Test anxiety in OT students was not found to be different between phases or time points. Spielberger's manual [26] states that the TAI was developed as a tool to measure test anxiety as a "situation-specific personality trait." This could potentially be less sensitive to changes over time than Form Y-1 of the STAI, which is a measure that is sensitive to changes in state anxiety. Note that test-retest reliability of TAI scores for college and graduate students for time spans of 3 weeks and 2 weeks, respectively, have each been found to be strong (r=.80) [26]. Personality traits are expected to be stable and unlikely to change over time [36]. Despite this lack of difference, there are two points to consider:

1. Having no significant difference in test anxiety scores between phases at the first time point, prior to the VR simulation intervention, further establishes similarity between phases of OT students for test anxiety–specific personality traits, prior to their shown differences in state anxiety at TP2.
2. The implication of test anxiety–specific personality traits being the same across the time points means that student attitudes toward clinical practical exams are unlikely to change as they participate in an OT program.

If students suffer debilitating symptoms of test anxiety, this is unlikely to change as they continue with their OT program. OT programs are recommended to be equipped with separate and dedicated activities to mitigate test anxiety in students.

In order for immersive VR to have potential in reducing trait-based anxiety, it would require an established treatment protocol, similar to what Gaggioli and colleagues stated in their workplace stress report in 2014 [20]. Their treatment protocol followed the stress-management training program as established by Kaluza [37] and Meichenbaum [38], which consisted of 10 1-hour sessions in 5 weeks, administered by clinical psychologists [37-39]. A stress-management training program utilizing VR in this manner would be expected to reduce chronic workplace trait anxiety by 12%, greater than the results found when compared to a cognitive behavioral therapy control group [20].

This investigation unexpectedly found academic self-efficacy to gradually increase across the time points for both phases, despite an expected moderate inverse relationship being found at the peak state anxiety point. Greater academic self-efficacy being associated with lower state anxiety is congruent to Dobson's report [8], which was reflected in this investigation's peak anxiety time point. In a report featuring anxiety related to writing in graduate students, self-efficacy was found to have a large and inverse relationship with writing anxiety [40]. However, it has been stated in previous reports that low-to-moderate levels of emotionality (ie, affective physiological reactions to anxiety) may actually enhance a student's performance, while excessive levels may cause a reduction in performance [41,42]. The peak state anxiety levels in the OT students within this investigation were not strong enough to show a noticeable reduction in their academic self-efficacy scores, possibly due to extraneous variables, such as mental resiliency and previous experience. Enjoyment in the learning material and student pride have been found to have positive associations for self-efficacy [43]. Future research that compares OT students' peak state anxiety levels and their actual OSCE performance scores would allow further conclusions to be made for academic self-efficacy, state anxiety, and performance relationships.

There is a conceivable argument to be made for the necessity of students to endure anxiety symptoms as they progress through OT programs. By overcoming situations that induce anxiety, it is arguable that students will learn necessary coping skills to apply in practice for the real world. However, this investigation showed no difference in the rate of academic self-efficacy development between the YesVR and NoVR groups of OT students, despite their significant difference in peak state anxiety scores at TP2. Thus, it is presumed that OT students will not have their academic self-efficacy development compromised when VR interventions significantly reduce state anxiety levels. VR interventions that are designed to reduce state anxiety do not result in OT students missing out on academic self-efficacy development.

For the association between total study time and peak state anxiety, it is to be assumed that greater time spent in preparation for an upcoming OSCE is equivalent to students placing greater amounts of perceived importance onto the successful outcome of the evaluation. Students who appraise exams with high importance are associated with increased state test anxiety levels before the exam, which results in higher anxiety levels after the exam [44]. This association could also mean that facilitative aspects of anxiety may have compelled the students to spend greater amounts of time in preparation for the OSCE.

Based on participant responses from the Previous Experience with VR Survey, it appears that views regarding the adoption of VR simulation into OT programs are favorable, especially for use in professional work and education. The majority of students stated that the quality of graphics was the most important consideration for achieving realistic virtual environments, while also stating that ease of use and interaction were also important. However, there were other important factors, besides graphics, which may be just as vital in order to achieve a higher sense of fidelity. Although the speech-recognition system allowed the use of natural language to create the clinical interview experience, there were no measures of participant frustration or concern whenever they asked a question that the virtual patient failed to understand. Although the participants were instructed to reword questions or change the subject when encountering such events, these limitations in speech-recognition software may have negatively impacted the user's sense of immersion, potentially impacting the system's anxiety-reduction capability. Future designs are recommended to include a measure of the amount of virtual patient communicative misunderstandings and user immersion levels, determining if these variables are related.

For the verbal communication factors, there were limitations that may have impacted the user experience. For example, the avatars in this investigation all spoke with the same tone. They did not change their tone of voice to signify differing levels of severity to their condition. Differing tones of voice could have been used to communicate various types of emotion, which may have added to a user's perceived level of fidelity. Depending on the choice of words a virtual patient uses, their messages could have negative or positive connotations. This investigation did not feature the changing of a virtual patient's spoken words to influence negative or positive connotations. For example, a virtual patient could have said either "I hurt my shoulder" or "I ruined my shoulder" to imply different messages. Future virtual patient dialogue would benefit from these verbal communication factors being considered.

For the nonverbal communication factors, the virtual standardized patients were programmed to maintain eye contact with the user throughout the simulation, yet this extent of nonverbal communication could be improved upon with developed posture, gesture, and facial expression. Kinesics, such as posture, gesture, and facial expression, are encouraged to be implemented into virtual avatars to optimize the quality of communicative experiences [45]. The communicative properties of an avatar's eyes (ie, oculesics), such as gaze, pupil dilation, and eyelid movements, are considered to have a major impact on a user's perceived sense of realism [45]. Based on Steptoe's report [45], varying these oculesics parameters in virtual standardized patients to match those of varying personality types or truth and deception responses may result in avatars that students may perceive to be socially real. A virtual environment that simulates having an interview with a patient, based mainly on social communication interactions, allows users to establish a sense of what is expected in the real world. Incorporating oculesics properties into virtual patients may instill a greater sense of immersion for the user to enhance their communicative experience.

Strengths and Limitations

This investigation is potentially the first to implement an immersive VR intervention within the discipline of OT for the reduction of state anxiety in students preparing for an OSCE. This investigation aimed to minimize researcher bias by having minimal contact with the participants. The primary outcome measures for state and test anxiety, in addition to academic self-efficacy, were taken by established theory-based tools. This investigation was supplemented by secondary outcome measures, including student total study time, study strategies, and previous experience with immersive VR, which were deduced from the primary measures.

However, this investigation had some limitations. Measures of student performance anxiety were not taken during the actual OSCE event, which did not allow inferences to be made about immersive VR's effectiveness for that specific occasion. There were no follow-up measures taken, such as during the students' next year of preparation for their second OSCE, to determine whether immersive VR had an effect on long-term memory development for *inhibitory learning*. Physiological stress markers, such as cortisol levels in the blood, saliva, and urine as well as heart rate, were not measured to determine possible changes in affective anxiety components. The sample size of this investigation was satisfied for within-subject measures, but the sample size would need to be increased for establishing greater confidence in the between-subject measures. This investigation was unable to check for student covariates between the phases and did not perform the rigors of a randomized controlled trial. The total cost for the software, hardware, and development of the simulation itself was estimated to be over US $50,000, which may discourage the adoption of such a system in other OT facilities. It is important to note that technological improvements in immersive VR hardware and software development are becoming increasingly efficient, resulting in an increase of accessibility to this platform.

Future Recommendations

In addition to implementing design changes for the rectification of limitations as stated in this investigation, future designs may consider the use of general or workplace-based self-efficacy questionnaires, establishing students' perceived levels of competency for the professional world. A comparison of immersive VR's performance, developed with improved artificial intelligence for interview skills training, evaluated with formal clinical performance assessments, could be implemented to establish students' levels of performance for the professional world. Virtual standardized patients could also be developed to have unique traits, allowing for students to train for scenarios that would otherwise be difficult or possibly dangerous.

Conclusions

This investigation shows evidence of immersive VR's capability to reduce anxiety in OT students who communicated with virtual standardized patients using natural language. Although test anxiety potentially leads to worry cognitions, which can disrupt students' attention, this investigation showed that academic self-efficacy continually increased in health science students as they persevered in their program. A combination of optimal study strategies and immersive VR simulation for the reduction of anxiety in health science students preparing for clinical practical exams can lead to a future of positive mental health change from the virtual to the real world. Will your next clinical interview take place in the virtual world?

Acknowledgments

The authors wish to acknowledge the contributions of the Rehabilitation Robotics Laboratory at the Edmonton Clinic Health Academy, which developed the OSCE simulation as described in this investigation. The study was supported by the Teaching and Learning Enhancement Fund (TLEF), University of Alberta.

Authors' Contributions

BJC, SE, and MRR co-conceptualized this investigation. SE contributed vision design of the clinical practical simulation. BJC and MRR outlined the psychometric tool selection, data intake, and analysis process. BJC led the manuscript writing process. SE and MRR contributed to the writing process and revisions. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey regarding previous experience with virtual reality technology.

[DOCX File , 17 KB - games_v8i3e18313_app1.docx]

Multimedia Appendix 2

Objective Structured Clinical Examination (OSCE) virtual reality avatar voice-recognition script (short version).

[DOCX File , 23 KB - games_v8i3e18313_app2.docx]

References

1. Grochowska A, Wichniak A, Jarema M. Virtual reality: A valuable tool to advance treatment of mental disorders. Arch Psychiatry Psychother 2019;21(1):65-73 [FREE Full text] [doi: 10.12740/app/101654]
2. Kniffin TC, Carlson CR, Ellzey A, Eisenlohr-Moul T, Beck KB, McDonald R, et al. Using virtual reality to explore self-regulation in high-risk settings. Trauma Violence Abuse 2014 Oct;15(4):310-321. [doi: 10.1177/1524838014521501] [Medline: 24874991]
3. Grenier S, Forget H, Bouchard S, Isere S, Belleville S, Potvin O, et al. Using virtual reality to improve the efficacy of cognitive-behavioral therapy (CBT) in the treatment of late-life anxiety: Preliminary recommendations for future research. Int Psychogeriatr 2014 Nov 10;27(7):1217-1225. [doi: 10.1017/s1041610214002300]
4. Bouras N, Holt G, editors. Psychiatric and Behavioural Disorders in Intellectual and Developmental Disabilities. Cambridge, UK: Cambridge University Press; 2007.
5. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. State-Trait Anxiety Inventory for Adults™ (Forms Y1 and Y2). Redwood City, CA: Mind Garden; 2015. URL: https://www.mindgarden.com/145-state-trait-anxiety-inventory-for-adults [accessed 2020-06-30]
6. Sommer M, Arendasy ME. Further evidence for the deficit account of the test anxiety–test performance relationship from a high-stakes admission testing setting. Intelligence 2015 Nov;53:72-80. [doi: 10.1016/j.intell.2015.08.007]
7. Liebert RM, Morris LW. Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychol Rep 1967 Jun;20(3):975-978. [doi: 10.2466/pr0.1967.20.3.975] [Medline: 6042522]
8. Dobson C. Effects of Academic Anxiety on the Performance of Students With and Without Learning Disabilities and How Students Can Cope With Anxiety at School [master's thesis]. Marquette, MI: Northern Michigan University; 2012 Apr 03. URL: https://www.nmu.edu/education/sites/DrupalEducation/files/UserFiles/Dobson_Cassie_MP.pdf [accessed 2020-06-30]
9. Gerwing TG, Rash JA, Allen Gerwing AM, Bramble B, Landine J. Perceptions and incidence of test anxiety. Can J Scholarsh Teach Learn 2015 Apr;6(3):1-17 [FREE Full text] [doi: 10.5206/cjsotl-rcacea.2015.3.3]
10. Vanstone DM, Hicks RE. Transitioning to university: Coping styles as mediators between adaptive-maladaptive perfectionism and test anxiety. Pers Individ Dif 2019 Apr;141:68-75. [doi: 10.1016/j.paid.2018.12.026]
11. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: WH Freeman; 1997.
12. Galyon CE, Blondin CA, Yaw JS, Nalls ML, Williams RL. The relationship of academic self-efficacy to class participation and exam performance. Soc Psychol Educ 2011 Dec 1;15(2):233-249. [doi: 10.1007/s11218-011-9175-x]
13. Zhang W. On college oral English teaching in the base of virtual reality technology. Appl Mech Mater 2014 Nov;687-691:2427-2430. [doi: 10.4028/www.scientific.net/amm.687-691.2427]
14. Bissonnette J, Dubé F, Provencher MD, Moreno Sala MT. Evolution of music performance anxiety and quality of performance during virtual reality exposure training. Virtual Real 2016 Jan 30;20(1):71-81. [doi: 10.1007/s10055-016-0283-y]
15. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. Behav Res Ther 2008 Jan;46(1):5-27. [doi: 10.1016/j.brat.2007.10.003] [Medline: 18005936]
16. Sotres-Bayon F, Cain CK, LeDoux JE. Brain mechanisms of fear extinction: Historical perspectives on the contribution of prefrontal cortex. Biol Psychiatry 2006 Aug 15;60(4):329-336. [doi: 10.1016/j.biopsych.2005.10.012] [Medline: 16412988]

17. Gutiérrez MA, Vexo F, Thalmann D. Stepping into Virtual Reality. London, UK: Springer; 2008.
18. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 2015;9:660 [FREE Full text] [doi: 10.3389/fnhum.2015.00660] [Medline: 26696869]
19. Rebelo F, Noriega P, Duarte E, Soares M. Using virtual reality to assess user experience. Hum Factors 2012 Dec;54(6):964-982. [doi: 10.1177/0018720812465006] [Medline: 23397807]
20. Gaggioli A, Pallavicini F, Morganti L, Serino S, Scaratti C, Briguglio M, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: A block randomized controlled trial. J Med Internet Res 2014 Jul 08;16(7):e167 [FREE Full text] [doi: 10.2196/jmir.3235] [Medline: 25004803]
21. Choiri MM, Basuki A, Bagus AY, Sukaridhoto S, Jannah M. Design and development virtual reality athletic: Virtual imagery to train sprinter's concentration. In: Proceedings of the 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC). 2017 Presented at: 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC); September 26-27, 2017; Surabaya, Indonesia p. 161-166. [doi: 10.1109/KCIC.2017.8228580]
22. Levac D, Glegg S, Colquhoun H, Miller P, Noubary F. Virtual reality and active videogame-based practice, learning needs, and preferences: A cross-Canada survey of physical therapists and occupational therapists. Games Health J 2017 Aug;6(4):217-228. [doi: 10.1089/g4h.2016.0089] [Medline: 28816511]
23. Bennett S, Rodger S, Fitzgerald C, Gibson L. Simulation in occupational therapy curricula: A literature review. Aust Occup Ther J 2017 Aug;64(4):314-327. [doi: 10.1111/1440-1630.12372] [Medline: 28378428]
24. Kirti R, Gopalakrishnan M. Objective Structured Clinical Examinations (OSCEs) as an assessment tool in undergraduate medical education. J Natl Med Assoc 2018 Dec;110(6):553-555. [doi: 10.1016/j.jnma.2018.02.001] [Medline: 30129500]
25. Maicher K, Danforth D, Price A, Zimmerman L, Wilcox B, Liston B, et al. Developing a conversational virtual standardized patient to enable students to practice history-taking skills. Simul Healthc 2017;12(2):124-131 [FREE Full text] [doi: 10.1097/sih.0000000000000195]
26. Spielberger CD, Gonzalez HP, Taylor CJ, Anton ED, Algaze B, Ross GR, et al. Test Anxiety Inventory. Sampler Set: Manual, Instrument, Scoring Guide. “Test Attitude Inventory”. Redwood City, CA: Mind Garden; 1980. URL: https://www.mindgarden.com/150-test-anxiety-inventory [accessed 2020-06-30]
27. Algaze B. Combination of Cognitive Therapy With Systematic Desensitization and Study Counseling in the Treatment of Test Anxiety [doctoral dissertation]. Tampa, FL: University of South Florida; 1980.
28. Gonzalez HP. The Effects of Three Treatment Approaches on Test Anxiety, Study Habits and Academic Performance [master's thesis]. Tampa, FL: University of South Florida; 1976.
29. Gonzalez HP. Effect of Systematic Desensitization, Study Counseling and Anxiety-Coping Training in the Treatment of Test Anxious Students With Good and Poor Study Habits [doctoral dissertation]. Tampa, FL: University of South Florida; 1987.
30. Jerusalem M, Mittag W. Selbstwirksamkeit, bezugsnormen, leistung und wohlbefinden in der schule [Self-efficacy, reference standards, performance and well-being at school]. In: Jerusalem M, Pekrun R, editors. Emotion, Motivation und Leistung [Emotion, Motivation and Performance]. Göttingen, Germany: Hogrefe Publishing Group; 1999:223-245.
31. Bandura A. On rectifying conceptual ecumenism. In: Maddux JE, editor. Self-Efficacy, Adaptation, and Adjustment: Theory, Research, and Application. New York, NY: Plenum Press; 1995:347-375.
32. Schwarzer R, Jerusalem M. Skalen zur Erfassung von Lehrer-und Schülermerkmalen: Dokumentation der Psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for Recording Teacher and Student Characteristics: Documentation of Psychometric Methods as Part of the Scientific Monitoring of the Model Experiment Self-Effective Schools]. Berlin, Germany: Freie Universität Berlin; 1999. URL: https://www.psyc.de/skalendoku.pdf [accessed 2020-06-30]
33. The definition of an OSCE. Medical Council of Canada. 2019. URL: https://mcc.ca/examinations/osce-orientation/definition/ [accessed 2019-09-09]
34. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
35. Downing SM, Haladyna TM. Validity threats: Overcoming interference with proposed interpretations of assessment data. Med Educ 2004 Mar;38(3):327-333. [doi: 10.1046/j.1365-2923.2004.01777.x] [Medline: 14996342]
36. Soldz S, Vaillant GE. The big five personality traits and the life course: A 45-year longitudinal study. J Res Pers 1999 Jun;33(2):208-232. [doi: 10.1006/jrpe.1999.2243]
37. Kaluza G. Changing unbalanced coping profiles: A prospective controlled intervention trial in worksite health promotion. Psychol Health 2000 May;15(3):423-433. [doi: 10.1080/08870440008402003]
38. Meichenbaum D. Stress Inoculation Training. New York, NY: Pergamon Press; 1985.
39. Pallavicini F, Gaggioli A, Raspelli S, Cipresso P, Serino S, Vigna C, et al. Interreality for the management and training of psychological stress: Study protocol for a randomized controlled trial. Trials 2013 Jun 28;14:191 [FREE Full text] [doi: 10.1186/1745-6215-14-191] [Medline: 23806013]
40. Huerta M, Goodson P, Beigi M, Chlup D. Graduate students as academic writers: Writing anxiety, self-efficacy and emotional intelligence. High Educ Res Dev 2016 Oct 04;36(4):716-729. [doi: 10.1080/07294360.2016.1238881]

41. Spielberger CD, Vagg PR. Test anxiety: A transactional process model. In: Spielberger CD, Vagg PR, editors. Test Anxiety: Theory, Assessment, and Treatment. Abingdon, UK: Taylor & Francis; 1995:3-14.
42. Cassady JC, Johnson RE. Cognitive test anxiety and academic performance. Contemp Educ Psychol 2002 Apr;27(2):270-295. [doi: 10.1006/ceps.2001.1094]
43. Villavicencio FT, Bernardo ABI. Beyond math anxiety: Positive emotions predict mathematics achievement, self-regulation, and self-efficacy. Asia Pac Educ Res 2015 Jul 23;25(3):415-422. [doi: 10.1007/s40299-015-0251-4]
44. Roick J, Ringeisen T. Self-efficacy, test anxiety, and academic success: A longitudinal validation. Int J Educ Res 2017;83:84-93. [doi: 10.1016/j.ijer.2016.12.006]
45. Steptoe WAH. Eye Tracking and Avatar-Mediated Communication in Immersive Collaborative Virtual Environments [doctoral dissertation]. London, UK: University College London; 2010 Sep 17. URL: https://pdfs.semanticscholar.org/538c/628c7ccfe3746619f671fe0b7cb6f0b33604.pdf?_ga=2.3948530.391761962.1593538024-245535511.1590015772 [accessed 2020-06-30]

Abbreviations

ANOVA: analysis of variance
ASE: Academic Self-Efficacy Scale
CPU: central processing unit
ηp_2: partial η_2
GPU: graphics processing unit
IPAT: Institute of Personality and Ability Testing
NoVR: subjects not exposed to the virtual reality simulation
OSCE: Objective Structured Clinical Examination
OT: occupational therapy
S-Anxiety: State Anxiety
STAI: State-Trait Anxiety Inventory
TAI: Test Attitude (Anxiety) Inventory
T-Anxiety: Trait Anxiety
TAS: Test Anxiety Scale
TLEF: Teaching and Learning Enhancement Fund
TMAS: Taylor Manifest Anxiety Scale
TP: time point
VR: virtual reality
VRET: virtual reality exposure therapy
WEQ: Worry Emotionality Questionnaire
YesVR: subjects exposed to the virtual reality simulation

Edited by G Eysenbach; submitted 18.02.20; peer-reviewed by S Jerdan, S Gallagher, F Alam; comments to author 10.03.20; revised version received 05.04.20; accepted 03.06.20; published 09.07.20.

Please cite as:
Concannon BJ, Esmail S, Roduta Roberts M
Immersive Virtual Reality for the Reduction of State Anxiety in Clinical Interview Exams: Prospective Cohort Study
JMIR Serious Games 2020;8(3):e18313
URL: http://games.jmir.org/2020/3/e18313/
doi:10.2196/18313
PMID:32673223

Original Paper

Reasons to Engage in and Learning Experiences From Different Play Strategies in a Web-Based Serious Game on Delirium for Medical Students: Mixed Methods Design

Kiki R Buijs-Spanjers[1,2], PhD; Harianne HM Hegge[1,2], MD; Fokie Cnossen[3], PhD; Debbie ADC Jaarsma[2], PhD; Sophia E de Rooij[1,4], MD, PhD

[1]Department of Geriatric Medicine, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

[2]Center for Education Development and Research in Health Professions, LEARN, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

[3]Bernoulli Institute of Mathematics, Computer Science and Artificial Intelligence, Department of Artificial Intelligence, University of Groningen, Groningen, Netherlands

[4]Medical Spectrum Twente, Medical School Twente, Enschede, Netherlands

Corresponding Author:
Kiki R Buijs-Spanjers, PhD
Department of Geriatric Medicine
University of Groningen
University Medical Center Groningen
Hanzeplein 1
P O Box 30.001 HP AA43
Groningen, 9700 RB
Netherlands
Phone: 31 503613464
Email: k.r.spanjers@umcg.nl

Abstract

Background: Although many studies have recently been published on the value of serious games for medical education, little attention has been given to the role of dark play (choosing unacceptable actions in games).

Objective: This study aimed to investigate potential differences in the characteristics of medical students who have the opportunity to choose normal or dark play in a serious game. This study also aimed to compare their reasons for choosing a play strategy and their perceptions of what they learned from their game play.

Methods: We asked undergraduate medical students to play a serious game in which they had to take care of a patient with delirium (The Delirium Experience). After getting acquainted with the game, students could opt for normal or dark play. Student characteristics (age, gender, experience with caring for older or delirious patients, and number of completed clerkships) were collected, and the Delirium Attitude Scale and Learning Motivation and Engagement Questionnaire were administered. Reasons for choosing normal or dark play were evaluated with an open-ended question. Information on lessons they had learned from the game was collected using an open-ended question and self-reported knowledge on delirium.

Results: This study had 160 participants (89 normal play, 71 dark play). Male students (26/160, 56.5%) chose dark play significantly more often than female students (45/160, 39.5%; *P*=.049). We did not find significant differences in student characteristics or measurement outcomes between play strategies. Participants' main reason for choosing normal play was to learn how to provide care to delirious patients, and the main reason for dark play was to gain insight into what a delirious patient has to endure during delirious episodes. All participants learned what to do when taking care of a delirious patient and gained insight into how a patient experiences delirium. We found no differences in self-reported knowledge.

Conclusions: When medical students have the opportunity to choose dark play in a serious game, half of them will probably choose this play strategy. Male students will more likely opt for dark play than female students. Choice of play strategy is not affected by any other student characteristic or measurement outcome. All students learned the same lessons from playing the game, irrespective of their learning strategy.

(JMIR Serious Games 2020;8(3):e18479) doi:10.2196/18479

KEYWORDS

dark play; serious games; medical education; medical students; delirium

Introduction

Players can engage in serious games with different play strategies. However, little is known on potential differences in characteristics of players in these different play strategies, their motivation to engage in that play strategy, or what they learn from it. In this study, we used a serious game on delirium to investigate these knowledge gaps.

Providing care for delirious patients poses a great burden on health care professionals [1,2]. Delirium is an acute neuropsychiatric syndrome that is characterized by altered attention, awareness, and cognition. Delirium is associated with longer hospital stays, functional decline, institutionalization, and mortality [3]. For patients, delirium also has negative effects on their psychological and emotional wellbeing [4,5]. Hence, it is important to understand delirious patients' needs when providing care to these patients, but, apparently, such understanding is often lacking [6,7]. Delirium often goes unrecognized by health professionals due to a lack of knowledge, awareness, and education about delirium. An overlap in symptoms of delirium and dementia makes it even harder to provide good quality care to delirious patients [1,8].

Current educational interventions mainly focus on knowledge and skills in recognizing delirium [9,10]. To improve delirium care, however, educational interventions need to have a different and broader focus. This includes gaining a better understanding of the patients' needs and health care professionals' attitudes towards delirious patients as well as promoting knowledge transfer to help learners develop knowledge of how to care for delirious patients [6,7]. It is important that educational interventions aimed at facilitating knowledge transfer encourage experiential learning in which learners are actively engaged with the study material [1]. In experiential learning, learners have to grasp and transform their experiences to create knowledge. In doing so, it is important that they are able to experiment with different approaches [11].

Serious games are interventions that promote experiential learning by providing a safe environment where learners can practice without the risk of harming the patient [12]. Serious games provide playful learning experiences that can be applied to real-life settings and actively involve learners [12]. They also give the learners autonomy on what they want to do, allowing them to experiment with different care options, which in turn will increase their feeling of control and satisfaction [13]. Moreover, serious games as experiential learning tools simultaneously allow learners to use different play strategies in the game (ie, normal or dark play). We define normal play strategy as choosing options in the game that resemble acceptable choices in real life. Another play strategy that players may opt for is dark play, during which players show in-game behaviors that are unacceptable in real life [14]. Experimenting with different types of care and options available in a serious game could provide learners with additional insights and knowledge [12,15].

Although many studies on serious games in medical education have been reported in recent years [16-18], these studies often investigated serious games as whole artifacts without focusing on the effect of different play strategies such as dark play. In a previous study, we showed that dark play did not affect game effectiveness [19]. However, our students had been allocated to a normal or dark play condition without being able to choose their game. Because of the allocation to a specific play strategy in the previous study, little is yet known about how often students voluntarily choose dark play in a serious game, which students choose to engage in dark play and why, and what they gain from their experience. By gaining insight on these aspects, we may meet the demand of more research on game elements that promote engagement and support the learning of players [16,18,20,21].

To be able to use dark play in serious games more efficiently in education and enhance learning by giving students an opportunity to experience the consequences of wrong choices and actions in a safe environment, more research on these topics is required. In this study, we sought to identify potential differences in characteristics between medical students who choose and do not choose to engage in dark play in a serious game on delirium. We examined their reasons for choosing normal or dark play and their perceptions of their learning experiences.

Methods

Educational Background

The master's program in medicine of the University Medical Center Groningen (UMCG) consists of 3 years. The first year is a dual learning year with 4 blocks, where each block consists of 5-week just-in-time skills training in a skills lab setting followed by 5-week "junior" clerkships. The second year comprises a series of ten 4-week "senior" clerkships, and the third year consists of a 20-week clinical elective and a 20-week research elective. Every 6 weeks, approximately 20 first-year master's students start their junior psychiatry clerkship. They play The Delirium Experience as part of their introductory program.

Participants, Recruitment, and Ethical Considerations

Participants in this study were first-year students of the master's program in medicine of the UMCG who were at the start of their psychiatry clerkship.

Between January 2018 and January 2019, at the start of each clerkship, all students were asked to participate in our study by the first author (KBS). They were informed about the purpose of the study, and afterwards they received digital information and a digital informed consent form. Participation was voluntary and could be stopped at any time. Students were also allowed to play the game without participating in the study. To ensure students did not feel obliged to participate, the researchers were not involved in other educational activities. All 160 students agreed to participate and signed the informed consent form.

Registration of the trial was not necessary in accordance with the International Committee of Medical Journal Editors recommendations.

Intervention

In our research, we used The Delirium Experience (Multimedia Appendix 1), a desktop simulation-based serious game that forces players to explore two different perspectives on delirium: that of a delirious patient (see Figure 1 for screenshots) and that of a health care professional (see Figure 2 for screenshots). The Delirium Experience was specifically developed to provide players with insight into what a delirious patient has to endure and how their actions and decisions as health care professionals may affect delirious patients and their interests [22].

The game works like this: during the daytime for 4 consecutive days, players take the role of a health care professional who provides care to a delirious patient. During the 4 nights, they switch to the patient's perspective to experience being delirious. Depending on the actions they choose during the day, their story will play out quite differently, and the delirious episodes will develop differently during the night. If players make the right care choices and provide good care, delirious episodes will be less severe than if they had made the wrong choices. Accordingly, the Delirium Experience enables players to opt for dark play by choosing wrong actions as a health professional and making the delirious episodes as severe as possible. Completing the entire game (all 4 days) takes about 20 minutes.

Players receive both direct and indirect feedback from the game. At the end of each day as a health care professional, they receive feedback on the consequences of their actions for the severity of the delirium and an overview of how their actions affected delirium severity. During the nights, players receive indirect feedback by experiencing the patient's responses to the actions of the health care professional during the delirious episodes.

Figure 1. Screenshots of the Delirium Experience serious game, from the patient's perspective.

Figure 2. Screenshots of the Delirium Experience serious game, from the health care professional's perspective.

Study Design and Procedure

We used triangulation during the interpretation of our findings to converge both quantitative and qualitative outcomes in a mixed-method design. We therefore concurrently gathered quantitative and qualitative data but analyzed these datasets separately.

Before playing The Delirium Experience, participants were asked to answer background questions about their age, gender, whether they had experience with caring for older and delirious patients (yes/no), and the number of clerkships they had completed. We also asked them to self-report their knowledge of delirium and to complete the Delirium Attitude Scale.

Participants played The Delirium Experience twice. The first time was meant to gain familiarity with the game. Before playing it a second time, students received a written explanation of the normal and dark play options in the game. When they played the game again, they had the opportunity to opt for another game play. They had the choice between playing the game with the intention to provide the best possible care to the patient (normal play; ie, guiding the patient through his hallucinations and/or supporting orientation) or to make the symptoms of delirium as severe as possible with their actions as a health care professional (dark play; ie, denying the patient's hallucinations and/or using sedation).

After playing, participants completed the Motivation and Engagement Questionnaire to evaluate their learning experiences. In addition, we asked them to indicate whether they had opted for normal or dark play, their reasons for choosing either normal or dark play, and what they had learned from it. Once again, we asked the participants to self-report their knowledge on delirium.

Outcome Measures

Self-reported knowledge on delirium was measured on a scale from 0 to 10, with higher scores indicating higher levels of perceived knowledge. In the Netherlands, it is common to use a mark like this that indicates your knowledge level. Marks ≥5.5 are considered sufficient knowledge, where marks <5.5 represent insufficient levels of knowledge.

Participants' attitudes towards delirious patients were measured using the Delirium Attitude Scale. This scale consists of 19 items that are scored on a 7-point Likert scale, with scores ranging from 19 to 133 points. Higher scores reflect a more positive attitude [23].

Participants' learning experiences were evaluated with the Motivation and Engagement Questionnaire, consisting of 9 items that are scored on a 5-point Likert scale, with scores ranging from 9 to 45 points [24]. Higher scores reflect more participant motivation and engagement in learning.

To uncover the reasons why participants chose normal or dark play, we added an open-ended question at the end of the digital questionnaire: "Why did you choose to play The Delirium Experience in normal or dark play?"

To investigate what participants had learned from their experience, we measured self-reported knowledge on delirium again (range 0-10) and asked the open-ended question: "What new insights did you gain while playing The Delirium Experience for the second time?"

Data Analysis

Quantitative Analysis

We checked data for normality by judging histograms, skewness, and kurtosis. To test for differences between participants who chose to engage in normal or dark play, we analyzed discrete variables (gender, experience with caring for older and delirious patients, and the number of clerkships they had completed) using chi-square tests and continuous variables (age, self-reported knowledge, attitude, and learning motivation and engagement) using independent samples *t* tests. *P* values <.05 were considered statistically significant. These statistical tests were performed with SPSS 23.0.

Qualitative Analysis

We thematically analyzed answers to the two open-ended questions ("Why did you choose to play The Delirium Experience in normal or dark play?" and "What new insights did you gain while playing The Delirium Experience for the second time?") with Atlas.ti software, version 8 (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany). We created separate data files to collect and analyze the responses to each open-ended question. During the coding process, the researchers were blinded to the participants' play strategy to ensure objectivity. After coding, we added information on the participants' play strategy to the data in order to analyze the differences.

We used inductive content analysis with constant comparison to find similarities and differences in students' answers between those who chose normal or dark play [25,26]. The whole process of qualitative data analysis was as follows. At first, the first author (KBS) read all answers to the open-ended questions to become familiar with the data. Subsequently, we identified initial codes, and KBS started coding the entire dataset. The resulting framework was iteratively refined as new, inductive codes were generated and integrated. Next, we identified preliminary themes by grouping similar concepts. Two researchers (KBS and DJ) reviewed and refined these preliminary themes to generate final themes. We analyzed the data in their original language; the most illustrative quotes were translated into English.

Results

Overview

In this section, we first describe the results of the baseline questions and reported learning motivation and engagement after playing The Delirium Experience, separately for students who chose normal or dark play. In the second part of this section, we describe students' self-reported reasons to engage in normal or dark play, and lastly, we report the results for students' self-reported knowledge gain in normal or dark play.

Participants Who Chose Normal or Dark Play

In total, 160 students participated in this study; 89 (56%) chose to play The Delirium Experience in normal play, and 71 (44%) chose dark play. Our study population consisted of 46 (46/160, 29%) male and 114 (114/160, 71%) female students, which is representative of the general student population of the master's program in medicine of the UMCG. Participants who chose normal or dark play did not significantly differ in age, experience with caring for older or delirious patients, or number of completed clerkships (Table 1). However, we found that male participants chose dark play significantly more often than female participants (56.5% [26/46] of men vs 39.5% [45/114] of women; *P*<.049). We did not find significant differences in self-reported knowledge on delirium, attitudes towards delirious patients before playing, or learning motivation and engagement after playing (Table 2).

Table 1. Characteristics of participants who chose normal or dark play.

Characteristics	Total (n=160)	Normal play (n=89)	Dark play (n=71)	*P* value
Age (years), mean (SD)[a]	23.0 (2.6)	23.6 (2.9)	23.1 (2.1)	.31
Gender, n (%)[b]				.049
Male	46 (28.8)	20 (43.5)	26 (56.5)	
Female	114 (71.2)	69 (60.5)	45 (39.5)	
Experience with older patients, n (%)[b]				.70
Yes	104 (65.0)	59 (56.7)	45 (43.3)	
No	56 (35.0)	30 (53.6)	26 (46.4)	
Experience with delirious patients, n (%)[b]				.78
Yes	56 (35.0)	32 (57.1)	24 (42.9)	
No	104 (65.0)	57 (54.8)	47 (45.2)	
Number of completed clerkships, n (%)[b]				.39
0	39 (24.4)	19 (48.7)	20 (51.3)	
1	38 (23.8)	23 (60.5)	15 (39.5)	
2	52 (32.5)	27 (51.9)	25 (48.1)	
3	28 (17.5)	17 (60.7)	11 (39.3)	
≥4	3 (1.9)	3 (100)	0 (0.0)	

[a]Data compared using independent samples *t* tests.

[b]Data compared using chi-square tests.

Table 2. Mean scores of self-reported knowledge, attitudes, and learning motivation and engagement in participants who chose normal or dark play.

Characteristic	Total (n=160)	Normal play (n=89)	Dark play (n=71)	*P* value[a]
Self-reported knowledge (possible score range, 0-10), mean (SD)	5.1 (1.9)	4.9 (1.9)	5.3 (1.8)	.23
Attitude (possible score range, 19-133), mean (SD)	90.8 (10.7)	91.1 (11.4)	90.4 (9.8)	.66
Learning motivation and engagement (possible score range, 9-45), mean (SD)	35.0 (4.0)	35.3 (4.2)	34.6 (3.9)	.35

[a]Data compared using independent samples *t* tests.

Reasons for Choosing Normal or Dark Play

Participants' reasons for choosing normal or dark play could be divided into three main themes: (1) to learn about delirium (care), (2) students' performance in the normal or dark game play, and (3) to take full advantage of the opportunities offered by the game.

To Learn About Delirium (Care)

A reason for participants to engage in normal play was that they considered learning how to provide good care for a delirious patient the most important and normal thing to do. One of the participants answered: "Because I feel it is more important [for me] to know how to act well." On the other hand, participants who had chosen dark play wanted to gain insight into what a delirious patient has to endure during delirious episodes: "I wanted to experience – from the patient's perspective – what it would be like to go through episodes of delirium."

Furthermore, participants were interested in seeing the progression of delirium. Participants who chose normal play wanted to see how delirium develops when providing good quality care and gain insight into factors that decrease the severity of delirious episodes. Participants who had chosen dark play, on the contrary, wanted to see how severe delirious episodes develop and which factors influence this.

Students' Performance in the Normal or Dark Game Play

A reason to choose either normal or dark play was that participants wanted to have a different game experience than they had in their first game. They thought it would be more instructive to see the effects of either correct or incorrect choices. For example, to explain why she used normal play, a participant answered: "During my first game play, I did not receive many points, and the delirium was quite severe, so I also wanted to see how delirium would progress if better treatment was given to the patient." A participant who had chosen dark play answered: "During my first game play, I became aware of what I could have done better. Therefore, I thought it would be more instructive to see [the stages of] progression of severe delirium."

In addition, participants who had chosen normal play wanted to see whether they had learned something from their first game play. They wanted to apply the knowledge they had obtained during their first game play and try to provide better care to the patient.

To Take Full Advantage of the Opportunities Offered by the Game

Another theme was the game itself and what its environment had to offer. In particular, participants who chose dark play indicated that the opportunities offered by the game environment was their reason to choose dark play: "In dark play, you can see what happens to a patient if you don't take good care of him; in real life, you just want to treat the patient as well as possible and [be able to] recognize the signs of poor treatment." These participants were also driven by curiosity about other scenarios in the game; as one participant said: "In the closing video, I saw some scenes with a doctor that I hadn't seen in the game yet."

Lessons Learned From Playing The Delirium Experience in Normal or Dark Play

To study lessons participants had learned from normal or dark play, we measured their self-reported knowledge on delirium and asked an open-ended question on what they had learned after playing. We did not find any differences in self-reported knowledge on delirium between participants who had chosen normal or dark play (mean 6.8, SD 1.2 vs mean 6.7, SD 1.2; t_{155}=0.361, *P*=.72).

Lessons participants had learned by playing The Delirium Experience for the second time can be divided into two themes: (1) an understanding of how to provide care to a delirious patient and (2) an understanding of the patient's experience. There were also participants who stated that they gained no new insights after playing The Delirium Experience for a second time.

An Understanding of How to Provide Care to a Delirious Patient

Participants' answers mainly focused on practical aspects of providing care to a delirious patient. First, participants saw the importance of guiding the patient, as stated by a participant who had chosen dark play: "The importance of good and correct contact with the patient, even though it seems hard due to the completely distracted [state of mind the] patient [was in]." Second, the importance of good orientation for the patient was frequently mentioned: "I experienced the game as really instructive and realized that small things, such as writing down the date and location, and opening the blinds, can contribute to decreased patient confusion." Third, participants gained new insights into prescribing medication for delirium: "I also need to realize that giving medication is not the most important thing to do."

Furthermore, participants gained more insights into how their actions as health care professionals could influence the patient and delirium: "As a health care professional, you are in control of how delirium develops, and you are able to worsen or improve it."

Finally, participants who had chosen normal as well as participants who had chosen dark play reflected on their knowledge while playing the game. A participant who had chosen dark play answered: "I thought I already had quite some knowledge of how to handle older people with delirium, but when the game forces you to make choices, this knowledge seemed to be limited."

An Understanding of the Patient's Experience

Participants who had chosen normal play as well as participants who had chosen dark play gained new insights into how a patient experiences delirious episodes. Participants answered:

> *I've never realized what it would be like to experience it as a patient, so this [playing The Delirium Experience] was really clarifying.*
>
> *The first time [I played The Delirium Experience], the delirium was not that exciting, but it was really scary for the patient the second time. It is good to see how frightening it can be.*

Furthermore, participants who chose dark play mentioned they had not expected that delirious episodes would be that intense.

No Added Value

Some of the participants who had chosen normal or dark play answered that there was not much added value in playing The Delirium Experience twice. The main explanations were that the game lacked feedback with reasons why they should act in a certain way and that the first game play was already instructive enough. For example, participants answered:

> *Little. I missed an explanation on why certain decisions were either positive or negative.*
>
> *Not really, the first time was more instructive.*

Discussion

Study Aims

With this study, we investigated potential differences in medical students who could choose between normal and dark play of a serious game and their perceived learning experiences. We therefore compared characteristics of students who opted for normal or dark play in a serious game on delirium. We investigated why students chose normal or dark play and what lessons they learned regarding normal or dark play.

Principal Findings

We found that male participants were more likely to choose dark play than female participants. We did not find any further differences between participants who chose normal or dark play in other characteristics (ie, age, experience with caring for older or delirious patients, and number of completed clerkships), attitude towards delirium, self-reported knowledge on delirium, and learning motivation and engagement. We grouped participants' reasons for choosing normal or dark play into three themes: (1) to learn about delirium (care), (2) students' performance in the normal or dark play game, and (3) to take full advantage of the opportunities offered by the game. The lessons participants learned after playing normal or dark play could be divided into two themes: (1) an understanding of how to provide care to a delirious patient and (2) an understanding of the patient's experience. We did not find any differences in self-reported knowledge after playing normal or dark play.

Our finding that male participants chose dark play more often than female participants may be explained by the following: In entertainment games, men are often player types who are more interested in exploring the game environment [27]. The results of our study suggest this may also be the case in serious games. Furthermore, there is evidence that men and women tend to have different task orientations. For example, women may be more interested in normal play because they tend to be more mastery-oriented than men, who tend to be more performance-oriented [28]. Interestingly, participants who chose normal play indicated that (one of) their reasons for choosing this type of game play was to learn how to provide care to delirious patients. Additionally, since women prefer entertainment games in which they have to make meaningful decisions [27,29], in serious games they may also be more interested in choosing a play strategy that includes making meaningful decisions (ie, providing good care). Lastly, because female medical students tend to score higher on empathy than their male peers [30], it may be harder for them to choose unacceptable gaming options that would harm a patient in real life.

The results of our study indicate that students learn the same lessons, irrespective of their learning strategy. Participants' reasons for choosing normal play centered around wanting to learn how to provide good quality care to delirious patients, while participants' reasons for choosing dark play pertained to the opportunity to experience what a delirious patient has to endure. Both groups of participants learned how to provide care to a delirious patient and gained insight into how a patient experiences delirious episodes. Although all participants played the game using different strategies and consequently experienced different simulations, they all seem to have learned the same lessons. A disadvantage of simulation-based education, however, may be the variety of situations that can occur in a simulation, which may result in different experiences and knowledge after the simulation [31]. Yet, our results imply that engaging in different simulation situations in The Delirium Experience (eg, the severity of the delirious episodes) does not result in differences in lessons learned or self-reported knowledge. This is in line with the results of our previous study showing that normal or dark play did not affect game effectiveness in students who were allocated to the two conditions [19].

Other authors have previously advocated that educational interventions on delirium should focus more on transfer of knowledge to practice [6,10]. Although we did not investigate providing care to a delirious patient in actual practice, participants did report many practical actions that can also help improve delirium care in practice. In addition, the aim of the Dutch Delirium Guidelines for health care professionals is to improve early recognition and treatment of delirium and delirium care [32]. It is therefore important that medical students are aware of the recommendations outlined in these guidelines to be able to provide good-quality delirium care. Many of the lessons our participants perceived to have learned during their game play were in line with the recommendations in the Dutch Delirium Guidelines (eg, guiding the patient and facilitating patient orientation).

Further Research

The feedback in the game was provided in two ways: directly at the end of each day in the game itself and in the form of the patient's response to the care choices made. Some participants indicated that playing the game using normal or dark play did not provide them with new insights, because they felt that feedback was lacking or unclear. One of the barriers for using feedback effectively may be students' inability to decode feedback [33]. Effects of feedback are strong when the feedback message is encouraging and specific [34]. To improve the effect of feedback in serious games, it may be interesting to study how players who claimed to have learned nothing new from playing the game received and recognized feedback during their game play. Players who do not recognize or understand the feedback will not be able to benefit from it.

Our study showed that students gained more insight into what a delirious patient endures during delirious episodes; however, in actual practice, understanding of the patient's needs is often lacking [6,7]. It would be interesting and relevant to study if and how students and health care professionals who work with delirious patients change their behaviors and attitudes when they encounter delirious patients in real life, after playing the game, particularly since destigmatization seems to occur when working closely with delirious patients who use to be stigmatized [35].

The demonstrated differences in female and male participants who chose normal or dark play are in line with the way male and female entertainment game players are categorized into player types [27,29]. However, research on gamification showed that design features can influence the preferences of player types [36]. Also, personality types and traits seem to play a role in which design features are most effective in serious games and gamification [37,38], which warrants further investigation into including personality traits when designing serious games for medical education. To develop tailored and effective serious games that match players' preferences, further research could be performed on specific preferences of serious games players, especially since disliked game elements can negatively affect outcomes and participation [36].

Conclusions

Serious games offer a safe environment for practicing real-life situations and for exploring options that are unacceptable in real life. Both types of game play can lead to the same learning outcomes. When students have an opportunity to play a serious game in dark play, almost half of the students will choose this type of game play. Male students are more likely to opt for dark play than their female peers. No other student characteristics influenced their choice of normal or dark play, nor did attitude, self-reported knowledge, or learning motivation and engagement. Irrespective of the strategy chosen, students reported the same lessons learned after playing a serious game on delirium in normal or dark play. They learned how to provide care to a delirious patient and gained insight into what a delirious patient endures.

Acknowledgments

We thank Tineke Bouwkamp-Timmer for editing the manuscript for language.

Conflicts of Interest

The serious game, Delirium Experience, was developed by IJsfontein and is owned by Stichting Effectieve Ouderenzorg (a Dutch foundation for improving elderly care by research and education). SEdR is an unpaid member of the supervisory board of Stichting Effectieve Ouderenzorg, which waived the licensing fee required for the use of intellectual property for the purposes of this research. The game is currently commercialized, but the revenues are solely used to improve current elderly care by gamification.

Multimedia Appendix 1
Trailer of The Delirium Experience.
[MP4 File (MP4 Video), 28721 KB - games_v8i3e18479_app1.mp4]

References

1. Brooke J, Manneh C. Caring for a patient with delirium in an acute hospital: The lived experience of cardiology, elderly care, renal, and respiratory nurses. Int J Nurs Pract 2018 Aug;24(4):e12643. [doi: 10.1111/ijn.12643] [Medline: 29532553]
2. Shipway DJH, Partridge JSL, Foxton CR, Modarai B, Gossage JA, Challacombe BJ, et al. Do surgical trainees believe they are adequately trained to manage the ageing population? A UK survey of knowledge and beliefs in surgical trainees. J Surg Educ 2015;72(4):641-647. [doi: 10.1016/j.jsurg.2015.01.019] [Medline: 25887505]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM). Washington, DC: American Psychiatric Association; 2013.
4. Weir E, O'Brien A. Don't go there - It's not a nice place: Older adults' experiences of delirium. Int J Ment Health Nurs 2019 Apr;28(2):582-591. [doi: 10.1111/inm.12563] [Medline: 30549214]
5. Instenes I, Gjengedal E, Eide LS, Kuiper KK, Ranhoff AH, Norekvål TM, CARDELIR Investigators. "Eight Days of Nightmares … " - Octogenarian Patients' Experiences of Postoperative Delirium after Transcatheter or Surgical Aortic Valve Replacement. Heart Lung Circ 2018 Feb;27(2):260-266. [doi: 10.1016/j.hlc.2017.02.012] [Medline: 28396186]
6. Fisher JM, Gordon AL, MacLullich AMJ, Tullo E, Davis DHJ, Blundell A, et al. Towards an understanding of why undergraduate teaching about delirium does not guarantee gold-standard practice--results from a UK national survey. Age Ageing 2015 Jan;44(1):166-170 [FREE Full text] [doi: 10.1093/ageing/afu154] [Medline: 25324329]
7. Teodorczuk A, Mukaetova-Ladinska E, Corbett S, Welfare M. Reconceptualizing models of delirium education: findings of a Grounded Theory study. Int Psychogeriatr 2013 Apr;25(4):645-655. [doi: 10.1017/S1041610212002074] [Medline: 23253310]
8. Morandi A, Davis D, Taylor JK, Bellelli G, Olofsson B, Kreisel S, et al. Consensus and variations in opinions on delirium care: a survey of European delirium specialists. Int Psychogeriatr 2013 Dec;25(12):2067-2075 [FREE Full text] [doi: 10.1017/S1041610213001415] [Medline: 23962713]
9. Coyle MA, Chang HC, Burns P, Traynor V. Impact of Interactive Education on Health Care Practitioners and Older Adults at Risk of Delirium: A Literature Review. J Gerontol Nurs 2018 Aug 01;44(8):41-48. [doi: 10.3928/00989134-20180626-02] [Medline: 30059138]
10. Yanamadala M, Wieland D, Heflin MT. Educational interventions to improve recognition of delirium: a systematic review. J Am Geriatr Soc 2013 Nov;61(11):1983-1993. [doi: 10.1111/jgs.12522] [Medline: 24219200]
11. Kolb D. Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice Hall; 1984.
12. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
13. Ryan R, Rigby C, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
14. Mortensen TE, Linderoth J, Brown AM, editors. The Dark Side of Game Play: Controversial Issues in Playful Environments. Abingdon, UK: Routledge; 2015.
15. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating Health Behavior Theory and Design Elements in Serious Games. JMIR Ment Health 2015;2(2):e11 [FREE Full text] [doi: 10.2196/mental.4133] [Medline: 26543916]
16. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]

17. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
18. Haoran G, Bazakidi E, Zary N. Serious Games in Health Professions Education: Review of Trends and Learning Efficacy. Yearb Med Inform 2019 Aug;28(1):240-248 [FREE Full text] [doi: 10.1055/s-0039-1677904] [Medline: 31022747]
19. Buijs-Spanjers KR, Hegge HH, Cnossen F, Hoogendoorn E, Jaarsma DA, de Rooij SE. Dark Play of Serious Games: Effectiveness and Features (G4HE2018). Games Health J 2019 Aug;8(4):301-306. [doi: 10.1089/g4h.2018.0126] [Medline: 30964340]
20. Boyle EA, Hainey T, Connolly TM, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education 2016 Mar;94:178-192. [doi: 10.1016/j.compedu.2015.11.003]
21. Wang R, DeMaria S, Goldberg A, Katz D. A Systematic Review of Serious Games in Training Health Care Professionals. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare 2016;11(1):41-51. [doi: 10.1097/sih.0000000000000118]
22. Effectieve Ouderen Zorg. Delirium Experience. 2017. URL: http://www.deliriumexperience.nl/ [accessed 2017-05-26]
23. Buijs-Spanjers KR, Hegge HH, Jansen CJ, Hoogendoorn E, de Rooij SE. A Web-Based Serious Game on Delirium as an Educational Intervention for Medical Students: Randomized Controlled Trial. JMIR Serious Games 2018 Oct 26;6(4):e17 [FREE Full text] [doi: 10.2196/games.9886] [Medline: 30368436]
24. Dankbaar MEW, Richters O, Kalkman CJ, Prins G, Ten Cate OTJ, van Merrienboer JJG, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. BMC Med Educ 2017 Feb 02;17(1):30 [FREE Full text] [doi: 10.1186/s12909-016-0836-5] [Medline: 28148296]
25. Bowling A, Ebrahim S. Handbook of Health Research Methods: Investigation, Measurement and Analysis. Berkshire: Open University Press; 2009.
26. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
27. Vahlo J, Kaakinen JK, Holm SK, Koponen A. Digital Game Dynamics Preferences and Player Types. J Comput-Mediat Comm 2017 Feb 08;22(2):88-103. [doi: 10.1111/jcc4.12181]
28. Kool A, Mainhard T, Brekelmans M, van Beukelen P, Jaarsma D. Goal orientations of health profession students throughout the undergraduate program: a multilevel study. BMC Med Educ 2016 Mar 31;16:100 [FREE Full text] [doi: 10.1186/s12909-016-0621-5] [Medline: 27036104]
29. Hartmann T, Klimmt C. Gender and Computer Games: Exploring Females' Dislikes. J Comp Mediated Comm 2006 Jul;11(4):910-931. [doi: 10.1111/j.1083-6101.2006.00301.x]
30. Hojat M, Gonnella JS. Eleven Years of Data on the Jefferson Scale of Empathy-Medical Student Version (JSE-S): Proxy Norm Data and Tentative Cutoff Scores. Med Princ Pract 2015 Apr 28;24(4):344-350 [FREE Full text] [doi: 10.1159/000381954] [Medline: 25924560]
31. Meyer M, Marzen-Groller K, Myers S, Busenhart C, Waugh S, Stegenga K. Simulation as a Learning Experience: Perceptions of New RNs. Clinical Simulation in Nursing 2014 Aug;10(8):384-394. [doi: 10.1016/j.ecns.2014.03.002]
32. Dautzenberg P, Molag M, van MB, de RS, Luijendijk H, Leentjens A. Herziene richtlijn Delier volwassenen en ouderen. Nederlands Tijdschrift voor Geneeskunde 2014;158.
33. Winstone N, Nash R, Rowntree J, Parker M. ‘It'd be useful, but I wouldn't use it’: barriers to university students’ feedback seeking and recipience. Studies in Higher Education 2016 Jan 18;42(11):2026-2041. [doi: 10.1080/03075079.2015.1130032]
34. van de Ridder JMM, McGaghie WC, Stokking KM, ten Cate OTJ. Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review. Med Educ 2015 Jul 15;49(7):658-673. [doi: 10.1111/medu.12744] [Medline: 26077214]
35. Mukaetova-Ladinska E, Cosker G, Chan M, Coppock M, Scully A, Kim S, et al. Delirium Stigma Among Healthcare Staff. Geriatrics (Basel) 2018 Dec 31;4(1):6 [FREE Full text] [doi: 10.3390/geriatrics4010006] [Medline: 31023974]
36. Kocadere S, Çaglar S. Gamification from Player Type Perspective: A Case Study. Educ Technol Soc Internet 2018;21(3):12-22 [FREE Full text]
37. Shabihi N, Taghiyareh F, Abdoli MH. Enhancement of educational games based on personality type indicators. International Journal of Information and Communication Technology Research 2017;9(3):37-45 [FREE Full text]
38. Orji R, Nacke L, Di MC. Towards Personality-driven Persuasive Health GamesGamified Systems. USA: ACM; 2017 Presented at: Proc 2017 CHI Conf Hum Factors Comput Syst Internet; 2017; New York, NY. [doi: 10.1145/3025453.3025577]

Abbreviations

UMCG: University Medical Center Groningen.

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by B Bonnechère, E Hoogendoorn, S Paige, H Verloo; comments to author 14.04.20; revised version received 14.05.20; accepted 20.05.20; published 29.07.20.

Please cite as:
Buijs-Spanjers KR, Hegge HHM, Cnossen F, Jaarsma DADC, de Rooij SE
Reasons to Engage in and Learning Experiences From Different Play Strategies in a Web-Based Serious Game on Delirium for Medical Students: Mixed Methods Design
JMIR Serious Games 2020;8(3):e18479
URL: http://games.jmir.org/2020/3/e18479/
doi:10.2196/18479
PMID:32723710

Original Paper

Biosensor Real-Time Affective Analytics in Virtual and Mixed Reality Medical Education Serious Games: Cohort Study

Panagiotis E Antoniou[1*], BSc, MSc, PhD; George Arfaras[1*], BSc, MSc; Niki Pandria[1*], BSc, MSc; Alkinoos Athanasiou[1*], MD, MSc, PhD; George Ntakakis[1*], BSc, MSc; Emmanouil Babatsikos[1*], BSc, MSc; Vasilis Nigdelis[1*], BA, MSc; Panagiotis Bamidis[1*], BSc, MSc, PhD

Lab of Medical Physics, The Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

[*]all authors contributed equally

Corresponding Author:
Panagiotis Bamidis, BSc, MSc, PhD
Lab of Medical Physics
The Medical School
Aristotle University of Thessaloniki
PO Box 376
Thessaloniki, 54124
Greece
Phone: 30 2310 999310
Email: bamidis@auth.gr

Abstract

Background: The role of emotion is crucial to the learning process, as it is linked to motivation, interest, and attention. Affective states are expressed in the brain and in overall biological activity. Biosignals, like heart rate (HR), electrodermal activity (EDA), and electroencephalography (EEG) are physiological expressions affected by emotional state. Analyzing these biosignal recordings can point to a person's emotional state. Contemporary medical education has progressed extensively towards diverse learning resources using virtual reality (VR) and mixed reality (MR) applications.

Objective: This paper aims to study the efficacy of wearable biosensors for affect detection in a learning process involving a serious game in the Microsoft HoloLens VR/MR platform.

Methods: A wearable array of sensors recording HR, EDA, and EEG signals was deployed during 2 educational activities conducted by 11 participants of diverse educational level (undergraduate, postgraduate, and specialist neurosurgeon doctors). The first scenario was a conventional virtual patient case used for establishing the personal biosignal baselines for the participant. The second was a case in a VR/MR environment regarding neuroanatomy. The affective measures that we recorded were EEG (theta/beta ratio and alpha rhythm), HR, and EDA.

Results: Results were recorded and aggregated across all 3 groups. Average EEG ratios of the virtual patient (VP) versus the MR serious game cases were recorded at 3.49 (SD 0.82) versus 3.23 (SD 0.94) for students, 2.59 (SD 0.96) versus 2.90 (SD 1.78) for neurosurgeons, and 2.33 (SD 0.26) versus 2.56 (SD 0.62) for postgraduate medical students. Average alpha rhythm of the VP versus the MR serious game cases were recorded at 7.77 (SD 1.62) μV versus 8.42 (SD 2.56) μV for students, 7.03 (SD 2.19) μV versus 7.15 (SD 1.86) μV for neurosurgeons, and 11.84 (SD 6.15) μV versus 9.55 (SD 3.12) μV for postgraduate medical students. Average HR of the VP versus the MR serious game cases were recorded at 87 (SD 13) versus 86 (SD 12) bpm for students, 81 (SD 7) versus 83 (SD 7) bpm for neurosurgeons, and 81 (SD 7) versus 77 (SD 6) bpm for postgraduate medical students. Average EDA of the VP versus the MR serious game cases were recorded at 1.198 (SD 1.467) μS versus 4.097 (SD 2.79) μS for students, 1.890 (SD 2.269) μS versus 5.407 (SD 5.391) μS for neurosurgeons, and 0.739 (SD 0.509) μS versus 2.498 (SD 1.72) μS for postgraduate medical students. The variations of these metrics have been correlated with existing theoretical interpretations regarding educationally relevant affective analytics, such as engagement and educational focus.

Conclusions: These results demonstrate that this novel sensor configuration can lead to credible affective state detection and can be used in platforms like intelligent tutoring systems for providing real-time, evidence-based, affective learning analytics using VR/MR-deployed medical education resources.

(JMIR Serious Games 2020;8(3):e17823) doi:10.2196/17823

KEYWORDS

virtual patients; affective learning; electroencephalography; medical education; virtual reality; wearable sensors; serious medical games

Introduction

Affective Learning

According to Bloom's taxonomy of learning domains, there are three main domains of learning, namely cognitive (thinking), affective (emotion/feeling), and psychomotor (physical/kinesthetic) [1]. Specifically, in the affective domain, learning objectives focus on the learner's interests, feelings, emotions, perceptions, attitudes, tones, aspirations, and degree of acceptance or rejection of instructional content [2]. Despite the fact that defining emotion is a rather daunting process, the term has been defined as "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" [3]. Furthermore, emotion has been defined from a psychological stance as a conscious representation of what individuals feel, whereas from a neuropsychological point of view, emotion "is seen as a set of coordinated responses that take place when an individual faces a personally salient situation" [4].

The role of affect (emotions) is considered crucial in the learning process, as well as in influencing learning itself, since it is linked to notions such as motivation, interest, and attention [5]. Earlier still, it was postulated [6] that learning most often takes place during an emotional episode; thus, the interaction of affect and learning may provide valuable insight in how people learn. Similarly, the relationship between learning and affective states is evident in other studies as well. In this way, a person's affective condition may systematically influence how they process new knowledge. It has been reported [7] that expert teachers are able to have a positive impact on their students' learning by recognizing and responding to their emotional states. Accordingly, attributes like curiosity, among other affective states, are identified as an indicator of motivation [8]. Such attributes constitute a driver for learning, being useful to motivated and affectively engaged learners in order to become more involved and to display less stress and anger [9-11], greater pleasure and involvement [10], and less boredom [11].

The well-established pleasure, arousal, dominance (PAD) psychological model of emotional states [12,13] represents all emotions deploying the above three dimensions on a scale from negative to positive values. Pleasure regards how pleasant (joy) or unpleasant (anger, fear) one feels about something. Arousal corresponds to how energized or bored one feels, and dominance refers to how dominant or submissive one feels. Even though the PAD model was originally configured with three components, the first two, pleasure and arousal, seem to have been used to a greater extent by researchers than dominance [14], mainly due to the fact that "all affective states arise from two fundamental neurophysiological systems, one related to valence (a pleasure–displeasure continuum) and the other to arousal" [15].

Sensor-Based Affect Recognitions

Biosignals such as heart rate (HR), blood volume pressure, external body temperature, electrodermal activity (EDA), and electroencephalography (EEG) are the physiological signals of the human body that drastically change during changes of emotional state [16]. Analyzing the change of physiological signal recordings can determine the emotional state of the human [17] by using devices that record and track the changes of the physiological signals, called biosensors [18,19]. Analyzing the data recordings from those devices for emotional state detection has been the research interest of recent studies. EDA and EEG biosignals have been used for the detection of stress during indoor mobility [20,21]. Additionally, EEG biosignal classification for emotion detection using the valence-arousal model of affect classification was also explored by our group [22].

Emotional Content and Brain Activation With Regards to Learning

Affective states, such as fear, anger, sadness, and joy, alter brain activity [23,24] and are associated with the neurophysiological interaction between cortical-based cognitive states and subcortical valence and arousal systems [15]. Today, ever-increasing neuroimaging data verify the importance of specific emotion-related brain regions, such as the orbitofrontal cortex, the dorsolateral prefrontal cortex, the cingulate gyrus, the hippocampus, the insula, the temporal regions, and the amygdala, in forming emotions and in the process of learning [25-32]. The hippocampus and amygdala form an apparatus of memory that houses two distinct yet functionally interacting mnemonic systems of declarative and nondeclarative memory [33,34]. Along with the other previously mentioned areas, they belong to the limbic system and the pathway of memory formation, consolidation, and learning [35], although those are wider processes that are not intrinsically tied to anatomical restraints and cannot be precisely localized [36,37].

Based on studies in animals and humans, which showed evidence of the critical role of the amygdaloid complex (AC) in emotional reactions [38-40], other studies in humans using functional magnetic resonance imaging have identified AC activation in response to affectively loaded visual stimuli [41-43]. Additionally, positron emission tomography scan studies have identified a connection between emotional stimuli and activity in the left AC [44]. The AC takes part in a system of nondeclarative memory formation and emotional conditioning using mechanisms such as long-term potentiation [45,46]. A functional asymmetry has also been identified between left and right AC, as negative emotional conditioning, especially based on fear, leads to a nondeclarative learning process, tracked predominantly in left AC [47]. Emotion valence and information encoding favor this asymmetry with regard to both positive and negative stimuli [48,49], while arousal has been linked to electroencephalographic theta waves from the amygdala [50]. This asymmetry also entails different modalities of encoded information and, while the left side has been functionally associated with language and detailed affective information, the right side has been functionally associated with imagery [49]. Moreover, it has been demonstrated that the declarative and nondeclarative systems interact during the process of learning, as emotion influences encoding by modulating the qualitative characteristics of attention, while episodic and active learning have also been proven to condition emotional response through memory formation [47,51-54].

Age difference is another factor that may potentially moderate cognitive appraisal of emotional content. Taking into consideration the age-related positivity effect, eye-tracking was used [55] to test for potential age differences in visual attention and emotional reactivity to positive and negative information between older and younger adults. It was discovered that when older adults processed negative emotional stimuli, they attended less to negative image content compared with younger adults, but they reacted with greater negative emotions. Finally, several studies have explored sex differences in the neural correlates of emotional content reactivity [41,56]. These have often highlighted the key role of the amygdala. It was found [57] that women exhibited increased activation in the amygdala, dorsal midbrain, and hippocampus. Similarly, men exhibited increased activation in the frontal pole, the anterior cingulate cortex and medial prefrontal cortex, and the mediodorsal nucleus of the thalamus.

Technology-Enhanced Immersive Medical Education and Virtual Reality

Information and communication technologies (ICT) have shaped interventions for health care and wellness from their beginning. Digital innovations reduce costs, increase capacities to support growth and address social inequalities, and improve diagnostic efficacy and treatment effectiveness. Contemporary medical education in particular has progressed extensively towards widely diverse learning resources and health care–specific educational activities in the ICT domain [58]. The incentive behind this lies with the necessity for worldwide access to clinical skills, unconstrained by time and place [59]. This potential of ICT in medical education is multiplied by the parallel advancement of web technologies and the proliferation of interactive learning environments with immediate, content-related feedback [60].

Currently, medical education is mostly based on case-based or problem-based learning and other small-group instructional models [61,62]. These include simulations, scenario narratives and other structured, task-based learning episodes. Scenario narratives in particular, termed virtual patients (VPs) in the health care sector, are serious game episodes designed custom to the learning objectives but also aligned with the expectations and skill sets of students in order to provide a game-informed, media-saturated learning environment. In that way, students can explore a case through multiple avenues, exercise their decision-making skills, and explore the impact of those decisions in a safe but engaging way [63,64]. VPs are defined as “interactive computer simulations of real-life clinical scenarios for the purpose of healthcare and medical training, education or assessment” [65]. Web-based VPs, unlike real patients, are consistently repeatable, since they are structured as branching narratives [66], offering few limitations with respect to time, place, and failure during the practice of clinical skills. Medical students have the opportunity to practice on a diverse set of rare and difficult diseases that they may later encounter in clinical practice [67]. Finally, the reproducibility of the case outcomes and the provisions for standardized validated assessment that exist in most VP platforms have established the use of VPs as an effective and important tool for modern medical education [68-70]. Due to these advantages, there is a global trend towards increased development of VPs, with many academic institutions working towards this goal [71]. The extended impact of VPs for medical education has been recognized, and standardization solutions for repurposing, reusing, and transferability have been initiated early on [65], with a formal standard, the MedBiquitous virtual patient standard, being finalized as early as 2010 [72,73]. Contemporary improvements, such as semantic annotations, have been implemented for easy reusability of VP content [74], while other efforts have focused on various fields, like elderly care [75], and even on intensifying experiential means, like virtual worlds [76,77], virtual reality, and augmented reality [78].

On the experiential front, many ideas have been implemented. One of them is the virtual laboratory. Virtual labs use simulations and computer models, along with a multitude of other media, such as video, to replace real-life laboratory interactions. A virtual lab consists of several digital simulations supported by discussion forums and video demonstrations, or even collaboration tools and stand-alone complex simulations [79]. Such interactive environments facilitate self-directed, self-paced learning (eg, repeating content, accessing content at off hours). That way, learners maintain initiative and increased engagement in the learning process, while interactivity hones laboratory skills that go beyond simple knowledge transfer. These laboratory skills strengthen the core areas of weakness in the contemporary medical curriculum. Hands-on laboratory techniques are usually not available for training to students due to cost, time, or safety constraints [80,81]. This leaves medical students with theoretical understanding but a lack of real-world clinical and lab skills [81].

An approach readily supportive of the virtual lab that can incorporate VP serious gaming is the implementation of virtual reality (VR), augmented reality, and, recently, mixed reality (MR) technologies, especially with the advent of devices like the Microsoft HoloLens (Microsoft Corp). The distinctions are somewhat blurred at times, but virtual reality is the substitution of external sensory inputs (mainly visual and audio) with computer-generated ones using a headset device. Augmented reality is the superposition of digital content over the real world that uses either 2-dimensional (2D) or 3-dimensional (3D) markers in the real-world environment. Finally, mixed reality is similar to augmented reality, with one key difference. Instead of the real-world marker being a preprogrammed static item or image, the superposition of content is done after a 3D mapping of the current environment has been completed. This way, features can be used in intuitive ways, like 3D models positioned on tables or 2D images or notes hanging on walls. There is evidence that such technologies significantly increase the educational impact of a learning episode and can subsequently greatly affect educational outcomes [82]. Realized examples include experiential world exploration [83], physics and chemistry concept visualizations with high engagement impact [84-86], and even the incorporation of such modalities for VPs [78]. It is this immediate engagement capacity of these modalities that can not only motivate the student but also allow for internalization of the educational material and, thus, avoidance of conceptual errors [87].

Aim and Scope of This Work

From this introduction, it becomes apparent that there is currently a sufficient body of research identifying both the impact and the capacity of digital tools for detecting and affecting the emotional state of users in their learning activities. Contemporary integrated wearable and unobtrusive sensor suites can provide objective, biosignal-based (as opposed to self-reported or inferred) emotion recognition. In addition, the proliferation and impact of immersive resources as medical education support tools provide strong motive to explore the feasibility of implementing in them real-time, evidence-based affective analytics.

Using commercial wearables and EEG sensors, this work presents the first, to the authors' knowledge, feasibility pilot of real-time, evidence-based affective analytics in a VR/MR-enhanced medical education VP serious game.

Methods

Equipment, Affective, and Educational Setup

A total of 11 healthy participants took part in our study after providing written informed consent. The participants included medical students (4 participants), medical school postgraduates (3 participants), and neurosurgeons (4 participants). The participants were informed that they would take part in 2 VP scenarios. The 2 scenarios were (1) a simple emergency response VP scenario that is familiar to all medical students beyond the third year of their studies, implemented on a simple web-based platform (OpenLabyrinth), and (2) a neuroanatomy-focused case regarding the ascending and descending pathways of the central nervous system, designed for a specialized neuroanatomy lecture at the graduate/postgraduate level and implemented in the MR HoloLens platform. The choice of scenario for the MR platform (Microsoft HoloLens) was forced to be this specific resource, as it was the only one that was available in scenario-based format. Both scenarios were medical narrative games that were guided by player choice. In the case of the web-based VP game, the users chose their responses from a multiple-choice panel on each of the VP game's pages. In the MR VP game, the users visualized the actual case through the HoloLens device, seeing and tacitly manipulating the relevant parts of the anatomy as they were described to them by narrative text. Selections in this modality were conducted by hand gestures in relevant locations of the actual anatomy that was presented to the user.

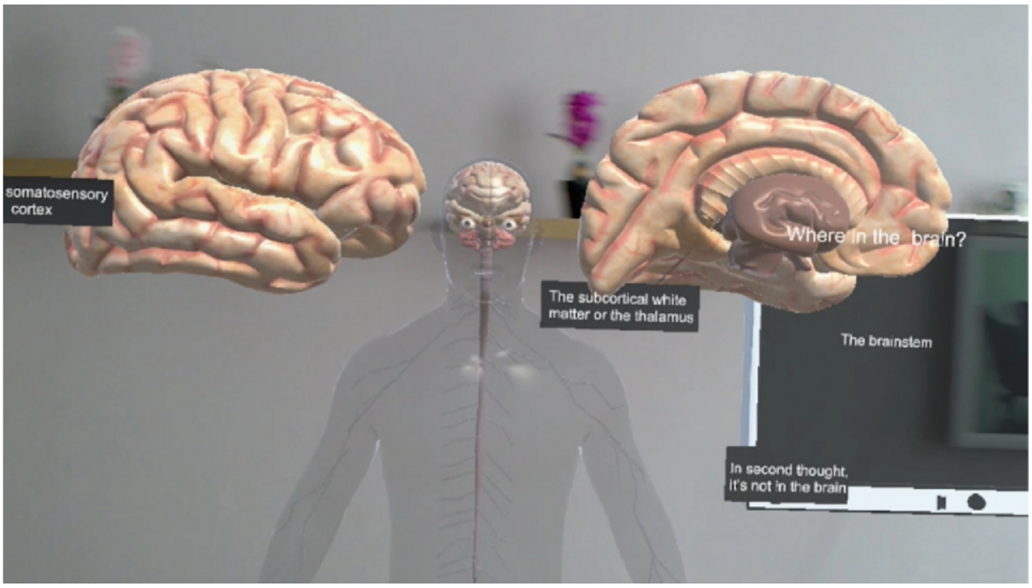
Due to the diverse background of the participants, they were exposed to the educational episodes in different ways. All groups tackled the first scenario on their own according to their knowledge. The medical students were fully guided in the MR scenario, since the educational content of it was beyond their knowledge. That means that they were free to ask any technical or medical question to the research team and they were provided with full guidance to select the correct answer. The postgraduate medical students were asked to resolve the MR scenario on their own and were offered assistance if they appeared to be stuck in and unable to proceed from a specific stage of the scenario. That means that they were free to ask any technical or medical question to the research team and they were provided with full guidance to select the correct answer. The neurosurgeons were all asked to complete the case on their own and were provided help only towards usage issues regarding the HoloLens device. That means that they were allowed to ask only technical questions about the functionality of the device.

In both scenarios, brain activity was recorded via an EEG, along with biosignals through a wearable unit. EEG signals were acquired using a 2-channel EEG amplifier (Nexus-10; Mind Media) [88] connected via Bluetooth to the PC, where signals were recorded (sampling rate 256 Hz) and preprocessed in real time. EEG electrodes were placed at the Fz and Cz positions, references at A1 and A2 (earlobes), and ground electrode at Fpz of the international 10-20 electrode placement system. Preprocessing included automatic EEG artifact removal and generation of real-time theta (4 to 8 Hz), alpha (8 to 12 Hz), and beta (13 to 21 Hz) rhythm. Moreover, the composite ratio of theta over beta was additionally generated.

For continuous, real-time physiological signals for stress detection, the E4 wearable multisensory smart wristband (Empatica Inc) was used [89]. HR and EDA were recorded with a sampling rate of 1 Hz and 4 Hz, respectively.

In order to implement the MR component of this experiment, the Microsoft HoloLens holographic computer headset was used [90]. Microsoft HoloLens is the world's first fully untethered holographic computer, providing holographic experiences in order to empower the user in novel ways. It blends optics and sensors to deliver seamless 3D content interaction with the real world. Advanced sensors capture information about what the user is doing, as well as the environment the user is in, allowing mapping and understanding of the physical places, spaces, and things around the user. The scenario used for the experiment was an exploratory interactive tutorial on the main central nervous system pathways in the brain and spinal cord (Figure 1). For the needs of this pilot experimental setup, 2 standard PC units were also used.

Figure 1. Part of the HoloAnatomy neuroanatomy virtual scenario.

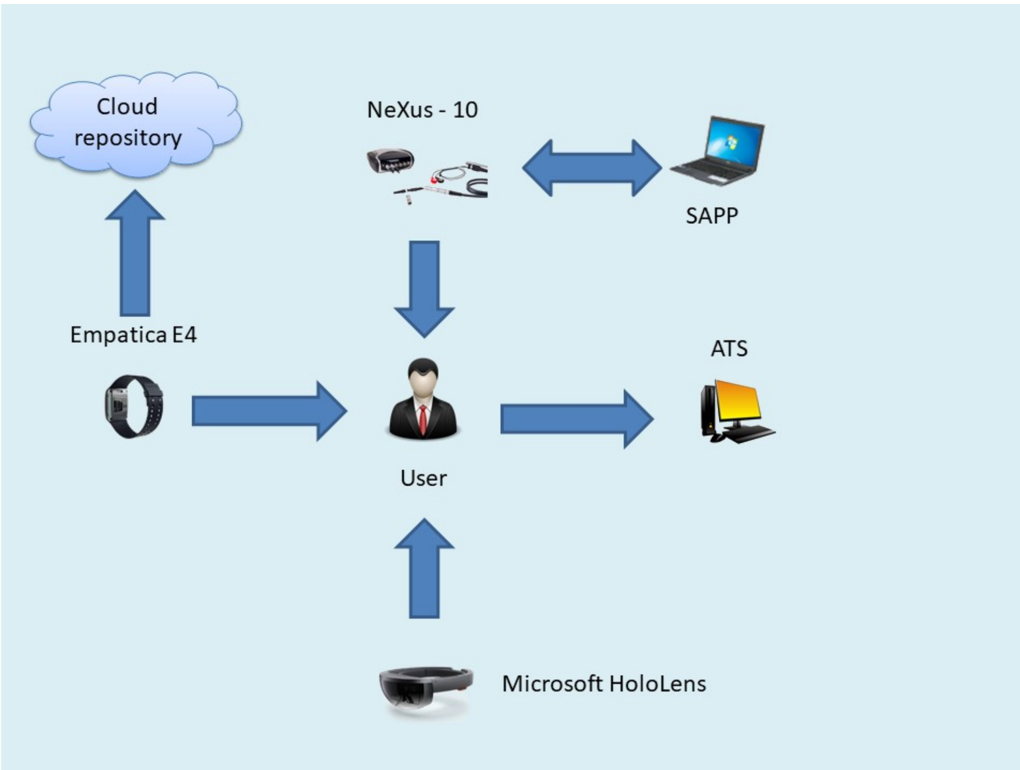


All real-time signal acquisition and postprocessing were conducted in a dedicated signal acquisition and postprocessing (SAPP) PC unit, while subjective emotion self-reporting, VP educational activity, and overall time stamp synchronization was conducted through an activity and time stamp synchronization (ATS) PC unit.

In the ATS unit, the Debut video capture software (NHS Software) [91] was used to record and time-stamp all of the participant's activities on screen for reference and manual synchronization with the internal clocks of the EEG and wearable sensor recorded from the SAPP unit.

The overall equipment setup and synchronization is demonstrated in Figure 2.

Figure 2. Multimodal biosignal sensors and mixed reality equipment setup. ATS: activity and time stamp synchronization; SAPP: signal acquisition and postprocessing.



Experimental Methodology

The experiment and relevant recordings took place on the premises of the Lab of Medical Physics of the Aristotle University of Thessaloniki in a quiet space. For the first session, the participant was invited to sit comfortably on a chair in front of the ATS unit's screen, located at a distance of 60 cm. While the participant was seated, the Empatica E4 sensor was provided to them to wear and the Nexus-10 EEG electrodes were applied. The ATS unit's screen displayed the initial page of the VP scenario and the time in Coordinated Universal Time (UTC) to allow for precise time recording. Data synchronization was conducted manually. The biosignal data acquired by the Empatica E4 and Nexus-10 devices were exported with a time stamp in UTC. Having a global time stamp both in the sensor recordings and in the educational activity computer (ATS) allowed for later offline synchronization. Specifically, the start of the time series for each sensor acquisition was synchronized in global time, together with the educational events as they were annotated manually, also in global time. The scenario is, in short, an interlinked web of pages that describes a coherent medical case.

The users navigated the interlinked web of pages by selecting their preferred answers in each part of the case from a predetermined multiple-choice list. The coordinator of the experiment was seated at a close distance next to the participant, though outside of the participant's visual field in order to not affect participant behavior, and was operating the SAPP unit continuously overseeing the acquisition process. The recordings of this session constituted the personalized sensor baseline for this user.

For the second session, the Empatica E4 and the Nexus-10 devices were used in the same manner as in the first session. Additionally, the HoloLens holographic computer unit was worn by the subject. Sitting comfortably on their chair, the participant viewed through the holographic unit an interactive exploratory neuroanatomy tour. Interaction with the HoloLens was conducted with gestures. In order not to contaminate the EEG recordings with the motor cortex EEG responses, the gestures were conducted by the coordinator after a preset time had passed (approximately 4 seconds). The whole session was video recorded in order to facilitate activity annotation at a later time. All participants took approximately the same time to finish their session. This time was approximately 35 to 40 minutes. About 10 minutes were used for orientation and equipment placement, 10 minutes were used for the VP case, and another 15 minutes were used for the MR experience.

Data Analysis

The acquired data (HR, EDA, alpha amplitude, theta/beta ratio) were annotated according to user activity data taken from the ATS unit and the video recording of the second session. Annotation marks in the web-based VP session were placed at the time points where the user moved to a new node in the VP scenario. These data segments of all the data sets (HR, EDA, alpha, theta/beta) formed the baseline values of each biosignal modality for this user. The acquired data of the MR VP session were annotated with marks placed at time points of the interaction gestures as they appeared in the video recording of the session. The first session segments were averaged in order to extract a global baseline average for each biosignal modality. The second session data were averaged on a per-segment basis, and the resulting data series (one data point per gesture per signal modality) were explored using descriptive statistics for quantitative differences from the baseline values.

Results

In the conventional educational episode, the participants explored a VP scenario. As previously mentioned, the data were annotated and segmented after each user transitioned from one stage of the VP scenario to the next. Thus, after averaging all biosignal data that were annotated and segmented for these stages, we extracted the averages for the biosignals recorded in this experimental setup. These included the alpha amplitude and the theta over beta ratio for the Cz EEG position (where the sensor was placed), as well as HR and EDA values. Representative results for one participant are summarized in Table 1.

A similar process was followed for the second session, where the VR/MR neuroanatomy resource was explored by the user. The time series of the biosignals were annotated and segmented on the time stamps corresponding to each gesture-based transition that the user experienced in this resource.

For each segment and for the HR and EDA, average values were recorded along with the segment number. Example plots are presented for a representative participant in Figure 3 and Figure 4. For reference purposes, the baseline average was also plotted as a constant in these graphs.

The same process was followed for the alpha rhythm and theta/beta ratio for the Cz point (sensor positions). Similar plots of the rhythms and ratios are presented for a representative participant in Figure 4 and Figure 5.

In Figure 4 and Figure 5, we have also included, as a plotted line, the average of the recorded signal in order to reveal even nondefinitive increases or decreases between the two experimental sessions. A representative value set is summarized in Table 2.

This analysis was performed for all 11 participants. The averaged results for each participant are presented in Table 3. Given the differentiation of each group's affective and educational setup, the participants are also partitioned according to group in this table. Table 4 presents a per-group average for all metrics recorded in the web-based VP and MR scenarios. Given that some of these metrics (eg, EDA) are highly varied across the population, which makes averaging irrelevant for any useful purpose (see standard deviations for EDA in Table 4), Table 5 presents the average value shifts between the two scenarios. Specifically, we present the total and per-group number of participants who had increased theta/beta ratios, decreased alpha rhythm amplitude, increased HR, and increased EDA in the MR scenario compared with the web-based VP scenario.

Table 1. Biosignal baseline averages of a representative participant (neurosurgeon) during the conventional educational episode.

Segment	Alpha, mean (SD), µV	Theta/beta power ratio, mean (SD)	HR[a], mean (SD), bpm	EDA[b], mean (SD), µS
Segment 1	4.886 (0.1822)	3.198 (0.5675)	86 (2)	0.227 (0.0114)
Segment 2	5.259 (0.1822)	2.235 (0.5675)	90 (2)	0.209 (0.0114)
Segment 3	4.883 (0.1822)	3.047 (0.5675)	88 (2)	0.203 (0.0114)
Segment 4	4.921 (0.1822)	3.584 (0.5675)	89 (2)	0.203 (0.0114)
Average	4.987 (0.1822)	3.016 (0.5675)	87 (2)	0.215 (0.0114)

[a]HR: heart rate.

[b]EDA: electrodermal activity.

Figure 3. Representative heart rate segment plot for the VR/MR educational episode. Dashed line denotes the baseline established in the first experimental session. HR: heart rate; MR: mixed reality; VR: virtual reality.

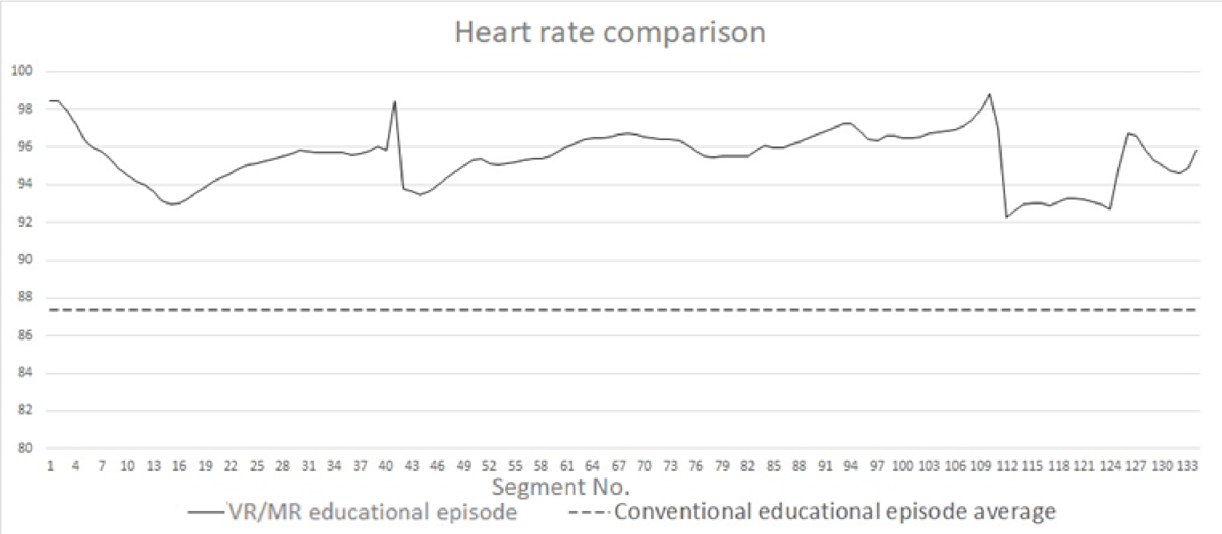


Figure 4. Alpha amplitude (µV) segment plot for the VR/MR educational episode for the Cz position. MR: mixed reality; VR: virtual reality.

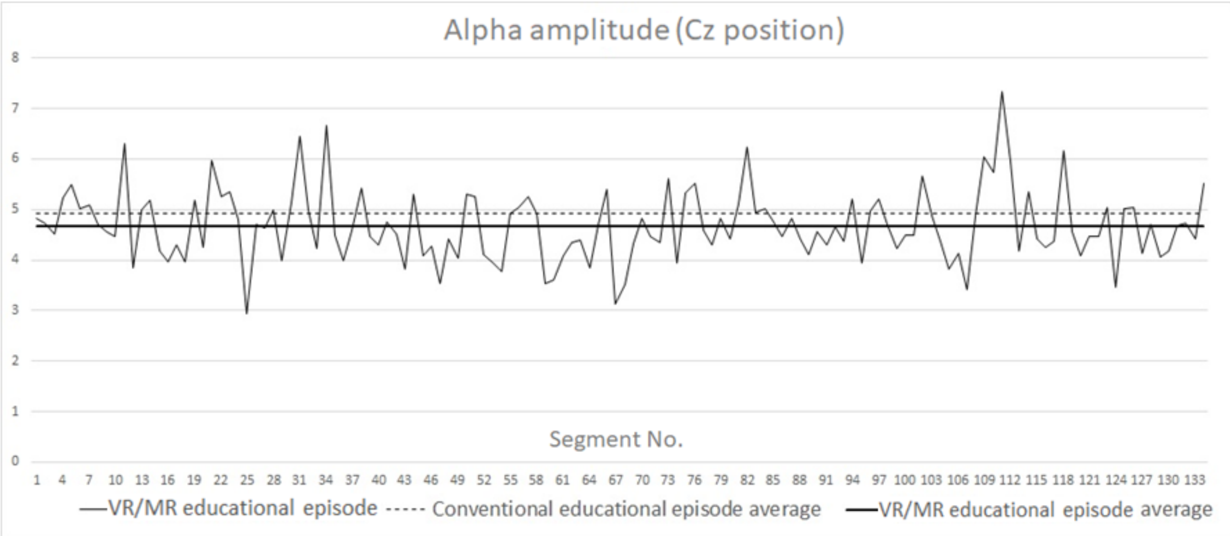


Figure 5. Theta/beta ratio segment plot for the VR/MR educational episode for the Cz position. MR: mixed reality; VR: virtual reality.

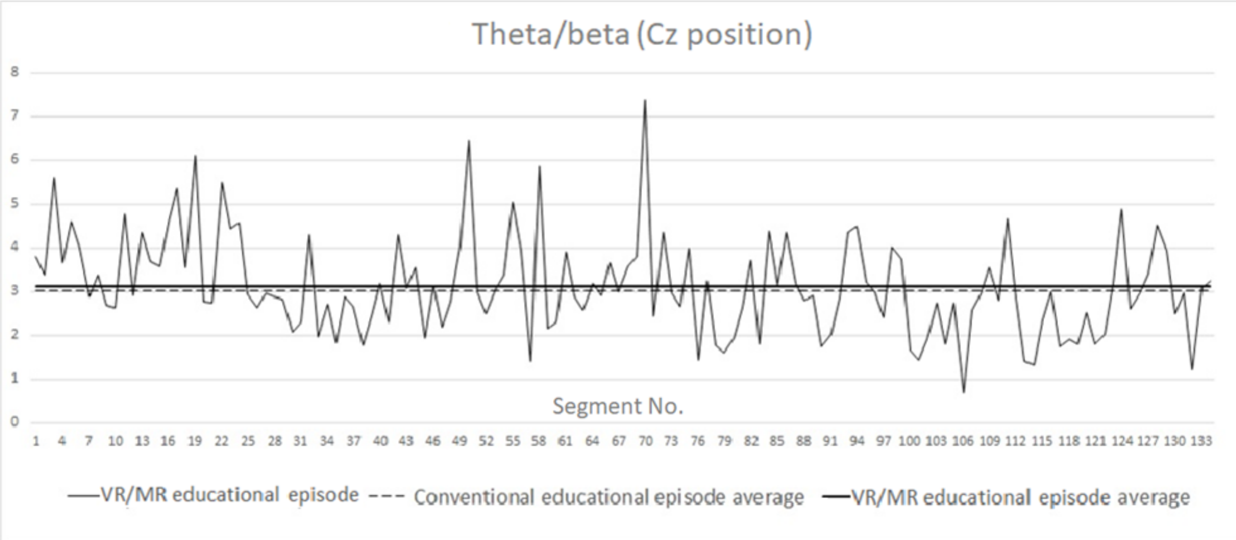


Table 2. Biosignal average of a representative participant (neurosurgeon) during the virtual reality/mixed reality educational episode.

	Alpha, mean (SD), µV	Theta/beta power ratio, mean (SD)	HR[a], mean (SD), bpm	EDA[b], mean (SD), µS
Average	4.680 (0.711)	3.131 (1.137)	95 (1)	3.475 (0.865)

[a]HR: heart rate.

[b]EDA: electrodermal activity.

Table 3. Biosignal averages per participant for the virtual reality/mixed reality educational episode.

Participant No.	Amplitude theta/beta (Cz), mean, µV		Amplitude alpha (Cz), mean, µV		HR[a], mean, bpm		EDA[b], mean, µS	
	VP[c]	MR[d]	VP	MR	VP	MR	VP	MR
Students								
1	4.315	4.417	8.024	8.540	97	96	3.733	6.479
2	2.734	1.788	10.228	12.443	92	92	0.525	2.438
3	4.306	3.425	5.794	5.529	66	65	0.268	0.382
4	2.611	3.284	7.040	7.157	96	90	0.265	7.090
Neurosurgeons								
1	3.016	3.131	7.033	7.146	87	95	0.215	3.475
2	1.438	1.598	6.023	6.223	88	82	5.788	14.539
3	1.986	1.679	6.396	8.072	75	76	1.019	3.076
4	3.935	5.420	10.725	9.606	73	79	0.538	0.540
Postgraduates								
1	2.439	2.532	7.831	7.801	75	70	0.371	0.307
2	2.580	3.331	20.523	13.933	90	84	0.387	4.516
3	1.971	1.823	7.163	6.926	77	79	1.459	2.671

[a]HR: heart rate.

[b]EDA: electrodermal activity.

[c]VP: virtual patient.

[d]MR: mixed reality.

Table 4. Biosignal averages per group for the virtual reality/mixed reality educational episode.

Group	Amplitude theta/beta power ratio (Cz), mean (SD)		Amplitude alpha (Cz), mean (SD), µV		HR[a], mean (SD), bpm		EDA[b], mean (SD), µS	
	VP[c]	MR[d]	VP	MR	VP	MR	VP	MR
Students	3.49 (0.82)	3.23 (0.94)	7.77 (1.62)	8.42 (2.56)	87 (13)	86 (12)	1.198 (1.467)	4.097 (2.79)
Neurosurgeons	2.59 (0.96)	2.90 (1.78)	7.03 (2.19)	7.15 (1.86)	81 (7)	83 (7)	1.890 (2.269)	5.407 (5.391)
Postgraduates	2.33 (0.26)	2.56 (0.62)	11.84 (6.15)	9.55 (3.12)	81 (7)	77 (6)	0.739 (0.509)	2.498 (1.72)

[a]HR: heart rate.
[b]EDA: electrodermal activity.
[c]VP: virtual patient.
[d]MR: mixed reality.

Table 5. Biosignal average value shifts for the virtual reality/mixed reality educational episode.

Group	Increased amplitude theta/beta power ratio (Cz), n	Decreased amplitude alpha (Cz), n, µV	Increased HR[a], n, bpm	Increased EDA[b], n, µS
Total	6	5	2	9
Students	2	1	0	4
Neurosurgeons	3	1	3	4
Postgraduates	2	3	0	2

[a]HR: heart rate.
[b]EDA: electrodermal activity.

Discussion

Originality

This work presented a pilot study for technically achieving the capacity to obtain evidence-based, real-time affective analytics from users of a VR/MR educational serious game resource. Three participant groups were included, namely undergraduate medical students, medical school postgraduates, and neurosurgeons. This is the first time that a multitude of recording and interaction devices were integrated within a cohesive and contemporary educational episode. This integration led to a plausible neurophysiological interpretation of recorded biosignals regarding engagement and other affective analytics metrics.

The technical barriers for this endeavor were significant. While the wrist-wearable sensor is unobtrusive and easy to wear, obtaining EEG recordings while simultaneously wearing a highly sophisticated electronic device like the Microsoft HoloLens is not easily completed. The literature has significant findings regarding EEG and VR, but usually these involve virtual environments (3D environments projected on a 2D screen) instead of true immersive virtual reality, which requires a dedicated headset [92]. In other cases, successful endeavors incorporating EEG to VR scenarios require expensive and cumbersome simulation rooms [93]. This work is the first, to the authors' knowledge, that involves (1) the first wearable, truly immersive mixed reality holographic computer (the MS HoloLens) and (2) real-time concurrent EEG recordings in an unconfined, free-roaming setting easily transferrable to real-world educational settings.

Principal Neurophysiological Results

Our neurophysiological investigation focuses primarily on the theta/beta ratio (the ratio of powers of theta rhythm to beta rhythm) and the amplitude of alpha rhythm (8 to 12 Hz). An increase of theta (4 to 8 Hz) power in EEG recordings has been documented, corresponding to neurophysiological processes that facilitate both working memory and episodic memory, as well as the encoding of new information [94,95]. Moreover, when recorded at the area over midline brain regions, theta activity is also related to cognitive processes that involve concentration, sustained attention, and creativity [96-100]. In line with this, higher theta activity has been reported in the frontal-midline regions during a task of high cognitive demand, a task with increasing working memory needs [101,102], or even a high-attention process [103-105]. On the other hand, engagement in attention-demanding tasks or judgment calls is reported to lead to alpha power suppression [106,107].

Our results, in the context of the educational and affective setup of the experiment, agree with the previously reported literature.

The undergraduate student group presented a decrease in theta/beta ratio activity and an increase in alpha rhythm amplitude. These results can be interpreted as engagement in high-demand cognitive functions while under suppression of judgement calls. As we described in the "Equipment, Affective, and Educational Setup" section, undergraduate students were offered the correct choices by the researchers during the MR scenario, given its very demanding neuroanatomy content. In that context, undergraduate students concentrated on the mechanical tasks of using the MR equipment, while no judgement calls were made by them (it can be said that cognitive control of the scenario was relegated to the researcher facilitating the student).

The second group, the postgraduate students, presented an increase in theta/beta ratio activity. They also presented a decrease in alpha rhythm amplitude during the MR serious game scenario. As we previously described, these students were asked to individually solve a very challenging (for their education level) medical scenario. Thus, they had to use all their cognitive faculties in order to overcome this challenge. The results are consistent with the educational setup for this group.

The third group, consisting of specialist neurosurgeon doctors, presented an increase in theta/beta ratio. They also presented an increase in alpha rhythm amplitude, similar to the undergraduate medical students. In this case, the participants (neurosurgeons) were presented with a scenario that required their concentration and initiative to solve, but they were not seriously challenged, since the educational material covered in the case was well within their skills. Thus, they had to commit cognitively to the task, especially to use the MR equipment, but not fully. This cognitive engagement ambivalence is demonstrated by the concurrent increase in theta/beta ratios and alpha rhythm amplitudes.

Regarding the EDA and HR results, these value were significantly elevated in the VR/MR session versus the baseline educational episode. Elevated HR and EDA are established signs of high arousal, independent of valence [108,109]. HR remained more or less steady on average in all groups, a fact that can be attributed to the overall relaxed environment of the experiment (seated participation, silent room, etc) and the resilience of the average HR in short-term variations. However, almost all the participants presented a significant increase in EDA, which can be attributed to the overall novelty factor of wearing a sense-altering digital device, as well as the immersion and excitement of the interaction with the VR/MR educational resource.

Limitations

Despite the overall promising results, this work contains some inherent limitations, mostly linked with the novelty of its aim and scope. A significant limitation is the small and diverse group of subjects that were used for the pilot run of the multimodal signal acquisition configuration. It must be emphasized that given the sampling rate of the biosensors, every 1 minute of recording provided 15,360 samples (256 Hz × 60 s) of EEG per channel, 60 samples (1 Hz × 60 s) of HR, and 240 samples (4 Hz × 60 s) of EDA. This high-density data throughout the study allowed for rather definitive biosignal results to be extracted on a per-participant basis. While the extensive data set gathered from participants provided credible results for this feasibility study, obviously there is the need to expand the participant sample in order to explore personalization and statistical verification challenges.

Another core limitation of this study is the low affective impact of the MR case. Furthermore, all patients experienced the educational content only once. The case had significant immersive content, including animation and audio cues as rewards and motivations for the user. However, it lacked impact and significant narrative consequences for the users' actions.

Thus, future work will require a larger user base, as well as more frequent exposure to emotional affective educational stimuli with emotional and narrative specificity of this suite in a VR/MR approach. Specific emotion-inducing content needs to be implemented in this modality in order to assess the biosignal variations as the emotional content clashes (or cooperates) with the VR/MR platform's inherently high arousal and emotional impact (for example, error-based training). Finally, as previously mentioned, a future goal to pursue based on this work is the integration of this biosensor suite in an intelligent tutoring system (ITS) for immediate medical teaching support that includes VR/MR resources.

Comparison With Prior Work

Similar endeavors for affective analytics in non–technology-heavy medical education episodes have already been initially explored [110]. However, this work is the first proof of application of at least one configuration that is realistically unconfined and applicable in a simple educational setting for emotion detection in a technology-heavy VR/MR-based medical education episode. The full scope of application for this work is the integration of sensors and devices for incorporating objective, sensor-based affective analytics in VR/MR educational resources provided by ITSs.

ITSs are computer systems that aim to provide immediate and customized instruction or feedback to learners [80], usually without requiring intervention from a human teacher. ITSs typically aim to replicate the demonstrated benefits of one-to-one (personalized) tutoring to one-to-many instruction from a single teacher (eg, classroom lectures) or no teacher at all (eg, online homework) [111]. An ITS implementation is based on students' characteristics and needs, and it analyzes and anticipates their affective responses and behaviors in order to allow more efficient collection of information of a student's performance, handle content adjustment, tailor tutoring instructions upon inferences on strengths and weaknesses, suggest additional work, and in general improve the learning level of students [112].

In that context, the integration of emotion detection for content and experience customization is not a new endeavor. Affective tutoring systems use bias-free physiological expressions of emotion (facial expressions, eye-tracking, EEG, biosensors, etc) in order "to detect and analyze the emotional state of a learner and respond in a way similar to human tutors" [113], hence being able to adapt to the affective state of students [114]. The results of this work provide initial evidence for the integration of multisensor configurations with a VR/MR platform and the synchronization and coordination of such a suite by an affect-aware ITS.

Conclusions

VR/MR is a versatile educational modality, especially in the highly sensitive medical domain. It provides the capacity both for highly impactful narrative, which is crucial for making doctors invested in the educational process, and for building decision-making skills and competences. It also provides the ability, through immersive simulation, for the medical learner to practice manual skills in surgical specialties. This dual capacity makes this modality highly impactful and sought after in the medical field. Therefore, the capacity to provide personalized, context-specific content and feedback based on the learner's emotional state is crucial in this modality for both self-directed and standard guided medical learning. This work provides the first proof of application for such endeavors.

Conflicts of Interest

None declared.

References

1. Bloom B. Bloom's taxonomy of learning domains. NWlink. URL: http://www.nwlink.com/~donclark/hrd/bloom.html [accessed 2020-03-09]
2. Krathwohl DR, Bloom BS, Masia BB. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York, NY: David McKay Co; 1964.
3. Scherer KR. Toward a Dynamic Theory of Emotion: The Component Process Model of Affective States. Geneva Studies in Emotion and Communication 1987;1:1-198 [FREE Full text]
4. Valenzi S, Islam T, Jurica P, Cichocki A. Individual Classification of Emotions Using EEG. JBiSE 2014;07(08):604-620. [doi: 10.4236/jbise.2014.78061]
5. Picard RW, Papert S, Bender W, Blumberg B, Breazeal C, Cavallo D, et al. Affective Learning — A Manifesto. BT Technology Journal 2004 Oct;22(4):253-269. [doi: 10.1023/b:bttj.0000047603.37042.33]
6. Stein NL, Levine LJ. Making sense out of emotion: The representation and use of goal-structured knowledge. In: Kessen W, Ortony A, Craik F, editors. Memories, thoughts, and emotions: essays in honor of George Mandler. New York: Psychology Press; 1991:322.
7. Goleman D. Emotional intelligence. New York, NY: Bantam Book Publishing; 1995.
8. Stipek DJ. Motivation to learn: from theory to practice. 3rd ed. Boston, MA: Allyn and Bacon; 1998.
9. Patrick BC, Skinner EA, Connell JP. What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. Journal of Personality and Social Psychology 1993;65(4):781-791. [doi: 10.1037/0022-3514.65.4.781]
10. Tobias S. Interest, Prior Knowledge, and Learning. Review of Educational Research 2016 Jun 30;64(1):37-54. [doi: 10.3102/00346543064001037]
11. Miserandino M. Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. Journal of Educational Psychology 1996;88(2):203-214. [doi: 10.1037/0022-0663.88.2.203]
12. Mehrabian J, Russell A. An approach to environmental psychology. Cambridge, MA: MIT Press; 1974.
13. Russell JA. A circumplex model of affect. Journal of Personality and Social Psychology 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
14. Bakker I, van der Voordt T, Vink P, de Boon J. Pleasure, Arousal, Dominance: Mehrabian and Russell revisited. Curr Psychol 2014 Jun 11;33(3):405-421. [doi: 10.1007/s12144-014-9219-4]
15. Posner J, Russel JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Develop Psychopathol 2005 Nov 1;17(03). [doi: 10.1017/s0954579405050340]
16. Chang C, Zheng J, Wang C. Based on Support Vector Regression for emotion recognition using physiological signals. 2010 Presented at: The 2010 International Joint Conference on Neural Networks (IJCNN); July 18-23, 2010; Barcelona, Spain URL: https://doi.org/10.1109/IJCNN.2010.5596878 [doi: 10.1109/IJCNN.2010.5596878]
17. Niu X, Chen Q. Research on genetic algorithm based on emotion recognition using physiological signals. 2011 Presented at: 2011 International Conference on Computational Problem-Solving (ICCP); Oct 21-23, 2011; Chengdu, China URL: https://doi.org/10.1109/ICCPS.2011.6092256 [doi: 10.1109/iccps.2011.6092256]
18. Takahashi K. Remarks on SVM-based emotion recognition from multi-modal bio-potential signals. 2004 Presented at: RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759); Sep 22, 2004; Kurashiki, Japan URL: https://doi.org/10.1109/ROMAN.2004.1374736 [doi: 10.1109/roman.2004.1374736]
19. Kim J, Bee N, Wagner J, Andre E. Emote to win: Affective interactions with a computer game agent. Lecture Notes in Informatics 2004;1:159-164.
20. Kalimeri K, Saitis C. Exploring multimodal biosignal features for stress detection during indoor mobility. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI. 2016 Presented at: 18th ACM International Conference on Multimodal Interaction - ICMI; Nov 12-16, 2016; Tokyo, Japan URL: https://doi.org/10.1145/2993148.2993159 [doi: 10.1145/2993148.2993159]
21. Ollander S, Godin C. A comparison of wearable and stationary sensors for stress detection. 2016 Presented at: IEEE International Conference on Systems, Man, and Cybernetics (SMC); Oct 9-12, 2016; Budapest, Hungary URL: https://doi.org/10.1109/SMC.2016.7844917 [doi: 10.1109/smc.2016.7844917]

22. Frantzidis C, Bratsas C, Klados M, Konstantinidis E, Lithari C, Vivas A, et al. On the Classification of Emotional Biosignals Evoked While Viewing Affective Pictures: An Integrated Data-Mining-Based Approach for Healthcare Applications. IEEE Trans Inform Technol Biomed 2010 Mar;14(2):309-318. [doi: 10.1109/titb.2009.2038481]
23. Lane RD, Reiman EM, Ahern GL, Schwartz GE, Davidson RJ. Neuroanatomical correlates of happiness, sadness, and disgust. Am J Psychiatry 1997 Jul;154(7):926-933. [doi: 10.1176/ajp.154.7.926] [Medline: 9210742]
24. Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat Neurosci 2000 Oct;3(10):1049-1056. [doi: 10.1038/79871] [Medline: 11017179]
25. Grodd W, Schneider F, Klose U, Nägele T. [Functional magnetic resonance tomography of psychological functions exemplified by experimentally induced emotions]. Radiologe 1995 Apr;35(4):283-289. [Medline: 7597162]
26. Schneider F, Gur RE, Mozley LH, Smith RJ, Mozley P, Censits DM, et al. Mood effects on limbic blood flow correlate with emotional self-rating: A PET study with oxygen-15 labeled water. Psychiatry Research: Neuroimaging 1995 Nov;61(4):265-283. [doi: 10.1016/0925-4927(95)02678-q]
27. Maddock RJ, Buonocore MH. Activation of left posterior cingulate gyrus by the auditory presentation of threat-related words: an fMRI study. Psychiatry Research: Neuroimaging 1997 Aug;75(1):1-14. [doi: 10.1016/s0925-4927(97)00018-8]
28. Canli T, Desmond JE, Zhao Z, Glover G, Gabrieli JDE. Hemispheric asymmetry for emotional stimuli detected with fMRI. Neuroreport 1998 Oct 05;9(14):3233-3239. [doi: 10.1097/00001756-199810050-00019] [Medline: 9831457]
29. Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, et al. Neural responses to facial and vocal expressions of fear and disgust. Proc Biol Sci 1998 Oct 07;265(1408):1809-1817 [FREE Full text] [doi: 10.1098/rspb.1998.0506] [Medline: 9802236]
30. Paradiso S, Johnson DL, Andreasen NC, O'Leary DS, Watkins GL, Ponto LL, et al. Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a PET study of normal subjects. Am J Psychiatry 1999 Oct;156(10):1618-1629. [doi: 10.1176/ajp.156.10.1618] [Medline: 10518175]
31. Lane RD, Reiman EM, Axelrod B, Yun L, Holmes A, Schwartz GE. Neural correlates of levels of emotional awareness. Evidence of an interaction between emotion and attention in the anterior cingulate cortex. J Cogn Neurosci 1998 Jul;10(4):525-535. [doi: 10.1162/089892998562924] [Medline: 9712681]
32. Wright CI, Fischer H, Whalen PJ, McInerney SC, Shin LM, Rauch SL. Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. Neuroreport 2001 Feb 12;12(2):379-383. [doi: 10.1097/00001756-200102120-00039] [Medline: 11209954]
33. Izquierdo I, Bevilaqua LRM, Rossato JI, Bonini JS, Silva WCD, Medina JH, et al. The connection between the hippocampal and the striatal memory systems of the brain: A review of recent findings. Neurotox Res 2006 Jun;10(2):113-121. [doi: 10.1007/bf03033240]
34. Klados M, Pandria N, Athanasiou A, Bamidis PD. A novel neurofeedback protocol for memory deficits induced by left temporal lobe injury [forthcoming]. International Journal of Bioelectromagnetism 2018.
35. Vann SD. Re-evaluating the role of the mammillary bodies in memory. Neuropsychologia 2010 Jul;48(8):2316-2327. [doi: 10.1016/j.neuropsychologia.2009.10.019] [Medline: 19879886]
36. Gaffan D. What is a memory system? Horel's critique revisited. Behav Brain Res 2001 Dec 14;127(1-2):5-11. [doi: 10.1016/s0166-4328(01)00360-6] [Medline: 11718881]
37. Horel JA. The neuroanatomy of amnesia. A critique of the hippocampal memory hypothesis. Brain 1978 Sep;101(3):403-445. [doi: 10.1093/brain/101.3.403] [Medline: 101278]
38. Adolphs R, Tranel D, Damasio H, Damasio A. Fear and the human amygdala. J Neurosci 1995 Sep 01;15(9):5879-5891. [doi: 10.1523/jneurosci.15-09-05879.1995]
39. Davis M. The role of the amygdala in conditioned fear. In: Aggleton JP, editor. The Amydala: Neurobiological Aspects of Emotion, Memory Mental Dysfunction. New York, NY: Wiley; 1992:395.
40. LeDoux JE. The neurobiology of emotion. In: Oakley DA, editor. Mind and Brain. Cambridge, MA: Cambridge University Press; 1986:354.
41. Reiman EM, Lane RD, Ahern GL, Schwartz GE, Davidson RJ, Friston KJ, et al. Neuroanatomical correlates of externally and internally generated human emotion. Am J Psychiatry 1997 Jul;154(7):918-925. [doi: 10.1176/ajp.154.7.918] [Medline: 9210741]
42. Irwin W, Davidson RJ, Lowe MJ, Mock BJ, Sorenson JA, Turski PA. Human amygdala activation detected with echo-planar functional magnetic resonance imaging. Neuroreport 1996 Jul 29;7(11):1765-1769. [doi: 10.1097/00001756-199607290-00014] [Medline: 8905660]
43. Müller JL, Sommer M, Wagner V, Lange K, Taschler H, Röder CH, et al. Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths. Biological Psychiatry 2003 Jul;54(2):152-162. [doi: 10.1016/s0006-3223(02)01749-3]
44. Taylor SF, Liberzon I, Fig LM, Decker LR, Minoshima S, Koeppe RA. The effect of emotional content on visual recognition memory: a PET activation study. Neuroimage 1998 Aug;8(2):188-197. [doi: 10.1006/nimg.1998.0356] [Medline: 9740761]
45. Phelps EA. Human emotion and memory: interactions of the amygdala and hippocampal complex. Curr Opin Neurobiol 2004 Apr;14(2):198-202. [doi: 10.1016/j.conb.2004.03.015] [Medline: 15082325]

46. Maren S. Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. Trends in Neurosciences 1999 Dec;22(12):561-567. [doi: 10.1016/s0166-2236(99)01465-4]
47. Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M. Activation of the left amygdala to a cognitive representation of fear. Nat Neurosci 2001 Apr;4(4):437-441. [doi: 10.1038/86110] [Medline: 11276236]
48. Lanteaume L, Khalfa S, Régis J, Marquis P, Chauvel P, Bartolomei F. Emotion induction after direct intracerebral stimulations of human amygdala. Cereb Cortex 2007 Jun;17(6):1307-1313. [doi: 10.1093/cercor/bhl041] [Medline: 16880223]
49. Markowitsch HJ. Differential contribution of right and left amygdala to affective information processing. Behav Neurol 1998;11(4):233-244 [FREE Full text] [doi: 10.1155/1999/180434] [Medline: 11568425]
50. Paré D, Collins DR, Pelletier JG. Amygdala oscillations and the consolidation of emotional memories. Trends Cogn Sci 2002 Jul 01;6(7):306-314. [doi: 10.1016/s1364-6613(02)01924-1] [Medline: 12110364]
51. Anderson AK, Phelps EA. Lesions of the human amygdala impair enhanced perception of emotionally salient events. Nature 2001 May 17;411(6835):305-309. [doi: 10.1038/35077083] [Medline: 11357132]
52. Fox E, Russo R, Bowles R, Dutton K. Do threatening stimuli draw or hold visual attention in subclinical anxiety? Journal of Experimental Psychology: General 2001;130(4):681-700. [doi: 10.1037/0096-3445.130.4.681]
53. Ochsner KN, Bunge SA, Gross JJ, Gabrieli JDE. Rethinking feelings: an FMRI study of the cognitive regulation of emotion. J Cogn Neurosci 2002 Nov 15;14(8):1215-1229. [doi: 10.1162/089892902760807212] [Medline: 12495527]
54. Funayama ES, Grillon C, Davis M, Phelps EA. A double dissociation in the affective modulation of startle in humans: effects of unilateral temporal lobectomy. J Cogn Neurosci 2001 Aug 15;13(6):721-729. [doi: 10.1162/08989290152541395] [Medline: 11564317]
55. Wirth M, Isaacowitz DM, Kunzmann U. Visual attention and emotional reactions to negative stimuli: The role of age and cognitive reappraisal. Psychol Aging 2017 Sep;32(6):543-556. [doi: 10.1037/pag0000188] [Medline: 28891668]
56. Lithari C, Frantzidis CA, Papadelis C, Vivas AB, Klados MA, Kourtidou-Papadeli C, et al. Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. Brain Topogr 2010 Mar;23(1):27-40 [FREE Full text] [doi: 10.1007/s10548-009-0130-5] [Medline: 20043199]
57. Filkowski MM, Olsen RM, Duda B, Wanger TJ, Sabatinelli D. Sex differences in emotional perception: Meta analysis of divergent activation. Neuroimage 2017 Feb 15;147:925-933. [doi: 10.1016/j.neuroimage.2016.12.016] [Medline: 27988321]
58. Fry H, Ketteridge S, Marshall S. A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice. New York, NY: Routledge; 2009.
59. Downes S. Distance Educators Before the River Styx Learning. URL: http://technologysource.org/article/ [accessed 2013-05-25]
60. Kaldoudi E, Konstantinidis S, Bamidis PD. Web 2.0 approaches for active, collaborative learning in medicine and health. In: Mohammed S, Fiaidhi J, editors. Ubiquitous Health and Medical Informatics: The Ubiquity 2.0 Trend and Beyond. Hershey, PA: IGI Global; 2010.
61. Williams B. Case based learning--a review of the literature: is there scope for this educational paradigm in prehospital education? Emerg Med J 2005 Aug;22(8):577-581 [FREE Full text] [doi: 10.1136/emj.2004.022707] [Medline: 16046764]
62. Larson RD. In Search of Synergy in Small Group Performance. New York, NY: Psychology Press; 2010.
63. Brief description.: ePBLnet Project URL: http://www.epblnet.eu/content/brief-description [accessed 2020-03-01]
64. Poulton T, Ellaway RH, Round J, Jivram T, Kavia S, Hilton S. Exploring the efficacy of replacing linear paper-based patient cases in problem-based learning with dynamic Web-based virtual patients: randomized controlled trial. J Med Internet Res 2014 Nov 05;16(11):e240 [FREE Full text] [doi: 10.2196/jmir.3748] [Medline: 25373314]
65. Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Medical Teacher 2009 Jul 03;30(2):170-174. [doi: 10.1080/01421590701874074]
66. Cook DA, Erwin PJ, Triola MM. Computerized Virtual Patients in Health Professions Education: A Systematic Review and Meta-Analysis. Academic Medicine 2010;85(10):1589-1602. [doi: 10.1097/acm.0b013e3181edfe13]
67. Saleh N. The Value of Virtual Patients in Medical Education. Ann Behav Sci Med Educ 2015 Oct 16;16(2):29-31. [doi: 10.1007/bf03355129]
68. Zary N, Johnson G, Boberg J, Fors UG. Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment--Web-SP. BMC Med Educ 2006 Feb 21;6:10 [FREE Full text] [doi: 10.1186/1472-6920-6-10] [Medline: 16504041]
69. Bateman J, Davies D. Virtual Patients: Are We in a New Era? Academic Medicine 2011;86(2):151. [doi: 10.1097/acm.0b013e3182041db4]
70. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Medical Teacher 2009 Sep 09;31(8):743-748. [doi: 10.1080/01421590903124708]
71. Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med 2007 May;82(5):446-451. [doi: 10.1097/ACM.0b013e31803e8a0a] [Medline: 17457063]
72. MedBiquitous Standards.: Medbuquitous URL: https://www.medbiq.org/medbiquitous_virtual_patient [accessed 2020-02-29]
73. Medbuquitous virtual patient summary.: Medbuquitous URL: https://www.medbiq.org/sites/default/files/vp_tech_overview/player.html [accessed 2020-03-01]

74. Dafli E, Antoniou PE, Ioannidis L, Dombros N, Topps D, Bamidis PD. Virtual patients on the semantic Web: a proof-of-application study. J Med Internet Res 2015 Jan 22;17(1):e16 [FREE Full text] [doi: 10.2196/jmir.3933] [Medline: 25616272]
75. Antoniou PE, Sidiropoulos EA, Bamidis PD. DISCOVER-ing beyond openSim; immersive learning for carers of the elderly in the VR/AR era. In: Communications in Computer and Information Science. 2017 Presented at: Third International Conference, Immersive Learning Research Network; June 26-29, 2017; Coimbra, Portugal URL: http://link.springer.com/10.1007/978-3-319-60633-0_16 [doi: 10.1007/978-3-319-60633-0_16]
76. Antoniou PE, Athanasopoulou CA, Dafli E, Bamidis PD. Exploring design requirements for repurposing dental virtual patients from the web to second life: a focus group study. J Med Internet Res 2014 Jun 13;16(6):e151 [FREE Full text] [doi: 10.2196/jmir.3343] [Medline: 24927470]
77. Antoniou PE, Ioannidis L, Bamidis PD. OSCase: Data Schemes, Architecture and Implementation details of Virtual Patient repurposing in Multi User Virtual Environments. EAI Endorsed Transactions on Future Intelligent Educational Environments 2016 Jun 27;2(6):151523 [FREE Full text] [doi: 10.4108/eai.27-6-2016.151523]
78. Antoniou PE, Dafli E, Arfaras G, Bamidis PD. Versatile mixed reality medical educational spaces; requirement analysis from expert users. Pers Ubiquit Comput 2017 Aug 23;21(6):1015-1024 [FREE Full text] [doi: 10.1007/s00779-017-1074-5]
79. Scheckler RK. Virtual labs: a substitute for traditional labs? Int J Dev Biol 2003;47(2-3):231-236 [FREE Full text] [Medline: 12705675]
80. de Jong T, Linn MC, Zacharia ZC. Physical and virtual laboratories in science and engineering education. Science 2013 Apr 19;340(6130):305-308. [doi: 10.1126/science.1230579] [Medline: 23599479]
81. Weller JM. Simulation in undergraduate medical education: bridging the gap between theory and practice. Med Educ 2004 Jan;38(1):32-38. [doi: 10.1111/j.1365-2923.2004.01739.x] [Medline: 14962024]
82. Chiu JL, DeJaegher CJ, Chao J. The effects of augmented virtual science laboratories on middle school students' understanding of gas properties. Computers & Education 2015 Jul;85:59-73. [doi: 10.1016/j.compedu.2015.02.007]
83. Dede C. Immersive interfaces for engagement and learning. Science 2009 Jan 02;323(5910):66-69. [doi: 10.1126/science.1167311] [Medline: 19119219]
84. Klopfer E, Squire K. Environmental Detectives—the development of an augmented reality platform for environmental simulations. Education Tech Research Dev 2007 Apr 5;56(2):203-228. [doi: 10.1007/s11423-007-9037-6]
85. Dunleavy M, Dede C, Mitchell R. Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. J Sci Educ Technol 2008 Sep 3;18(1):7-22. [doi: 10.1007/s10956-008-9119-1]
86. Wu H, Lee SW, Chang H, Liang J. Current status, opportunities and challenges of augmented reality in education. Computers & Education 2013 Mar;62:41-49. [doi: 10.1016/j.compedu.2012.10.024]
87. Olympiou G, Zacharia ZC. Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. Sci Ed 2011 Dec 20;96(1):21-47. [doi: 10.1002/sce.20463]
88. NeXus-10 product page.: Mind Media URL: https://www.mindmedia.com/en/products/nexus-10-mkii/, [accessed 2020-03-01]
89. Empatica E4 product page. : Empatica URL: https://www.empatica.com/en-eu/research/e4/ [accessed 2020-03-01]
90. Microsoft HoloLens product page.: Microsoft URL: https://www.microsoft.com/en-us/hololens [accessed 2020-03-01]
91. Debut Video Capture Software product page. : NCH Software URL: http://www.nchsoftware.com/capture/index.html [accessed 2020-03-01]
92. Charalambous EF, Hanna SE, Penn AL. Visibility analysis, spatial experience and EEG recordings in virtual reality environments: The experience of 'knowing where one is' and isovist properties as a means to assess the related brain activity. In: Proceedings of the 11th Space Syntax Symposium. 2017 Presented at: The 11th Space Syntax Symposium; July 3-7, 2017; Lisbon, Portugal p. 128.
93. Chin-Teng Lin, I-Fang Chung, Li-Wei Ko, Yu-Chieh Chen, Sheng-Fu Liang, Jeng-Ren Duann. EEG-Based Assessment of Driver Cognitive Responses in a Dynamic Virtual-Reality Driving Environment. IEEE Trans Biomed Eng 2007 Jul;54(7):1349-1352. [doi: 10.1109/tbme.2007.891164]
94. Karrasch M, Laine M, Rapinoja P, Krause CM. Effects of normal aging on event-related desynchronization/synchronization during a memory task in humans. Neurosci Lett 2004 Aug 05;366(1):18-23. [doi: 10.1016/j.neulet.2004.05.010] [Medline: 15265582]
95. Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 1999 Apr;29(2-3):169-195. [doi: 10.1016/s0165-0173(98)00056-3]
96. Kubota Y, Sato W, Toichi M, Murai T, Okada T, Hayashi A, et al. Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. Cognitive Brain Research 2001 Apr;11(2):281-287. [doi: 10.1016/s0926-6410(00)00086-0]
97. Başar-Eroglu C, Başar E, Demiralp T, Schürmann M. P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. International Journal of Psychophysiology 1992 Sep;13(2):161-179. [doi: 10.1016/0167-8760(92)90055-g]
98. Missonnier P, Deiber M, Gold G, Millet P, Gex-Fabry Pun M, Fazio-Costa L, et al. Frontal theta event-related synchronization: comparison of directed attention and working memory load effects. J Neural Transm (Vienna) 2006 Oct;113(10):1477-1486. [doi: 10.1007/s00702-005-0443-9] [Medline: 16604309]

99. Lagopoulos J, Xu J, Rasmussen I, Vik A, Malhi GS, Eliassen CF, et al. Increased theta and alpha EEG activity during nondirective meditation. J Altern Complement Med 2009 Nov;15(11):1187-1192. [doi: 10.1089/acm.2009.0113] [Medline: 19922249]
100. Gruzelier J. A theory of alpha/theta neurofeedback, creative performance enhancement, long distance functional connectivity and psychological integration. Cogn Process 2009 Feb;10 Suppl 1:S101-S109. [doi: 10.1007/s10339-008-0248-5] [Medline: 19082646]
101. Jensen O, Tesche CD. Frontal theta activity in humans increases with memory load in a working memory task. Eur J Neurosci 2002 Apr;15(8):1395-1399. [doi: 10.1046/j.1460-9568.2002.01975.x] [Medline: 11994134]
102. Grunwald M, Weiss T, Krause W, Beyer L, Rost R, Gutberlet I, et al. Theta power in the EEG of humans during ongoing processing in a haptic object recognition task. Cognitive Brain Research 2001 Mar;11(1):33-37. [doi: 10.1016/s0926-6410(00)00061-6]
103. Gevins A, Zeitlin G, Yingling C, Doyle J, Dedon M, Schaffer R, et al. EEG patterns during 'cognitive' tasks. I. Methodology and analysis of complex behaviors. Electroencephalography and Clinical Neurophysiology 1979 Dec;47(6):693-703. [doi: 10.1016/0013-4694(79)90296-7]
104. Gevins A, Zeitlin G, Doyle J, Schaffer R, Callaway E. EEG patterns during 'cognitive' tasks. II. Analysis of controlled tasks. Electroencephalography and Clinical Neurophysiology 1979 Dec;47(6):704-710. [doi: 10.1016/0013-4694(79)90297-9]
105. Gevins A, Zeitlin G, Doyle J, Yingling C, Schaffer R, Callaway E, et al. Electroencephalogram correlates of higher cortical functions. Science 1979 Feb 16;203(4381):665-668. [doi: 10.1126/science.760212] [Medline: 760212]
106. Adrian ED, Matthews BH. The interpretation of potential waves in the cortex. J Physiol 1934 Jul 31;81(4):440-471 [FREE Full text] [doi: 10.1113/jphysiol.1934.sp003147] [Medline: 16994555]
107. Niedermeyer E, Lopes da Silva F. Electroencephalography: Basic principles, clinical applications, and related fields. Baltimore, MD: Williams & Wilkins; 2004.
108. Fowles DC. The three arousal model: implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. Psychophysiology 1980 Mar;17(2):87-104. [doi: 10.1111/j.1469-8986.1980.tb00117.x] [Medline: 6103567]
109. Drachen A, Nacke L, Yannakakis G, Pedersen A. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games. 2010 Presented at: 5th ACM SIGGRAPH Symposium on Video Games; July 28, 2010; Los Angeles, CA p. 49-54. [doi: 10.1145/1836135.1836143]
110. Antoniou P, Spachos D, Kartsidis P, Konstantinidis E, Bamidis P. Towards Classroom Affective Analytics. Validating an Affective State Self-reporting tool for the medical classroom. MedEdPublish 2017;6(3) [FREE Full text] [doi: 10.15694/mep.2017.000134]
111. Blandford A. Intelligent tutoring systems: Lessons learned. Computers & Education 1990 Jan;14(6):544-545. [doi: 10.1016/0360-1315(90)90114-m]
112. VanLehn K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist 2011 Oct;46(4):197-221. [doi: 10.1080/00461520.2011.611369]
113. Mohanan R, Stringfellow C, Gupta D. An emotionally intelligent tutoring system. 2017 Presented at: 2017 International Conference on Distributed Computing Systems; July 18-20, 2017; London, England. [doi: 10.1109/sai.2017.8252228]
114. Sarrafzadeh A, Alexander S, Dadgostar F, Fan C, Bigdeli A. "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. Computers in Human Behavior 2008 Jul;24(4):1342-1363. [doi: 10.1016/j.chb.2007.07.008]

Abbreviations

2D: 2-dimensional
3D: 3-dimensional
AC: amygdaloid complex
ATS: activity and time stamp synchronization
EDA: electrodermal activity
EEG: electroencephalography
HR: heart rate
ICT: information and communication technologies
ITS: intelligent tutoring system
MR: mixed reality
PAD: pleasure, arousal, dominance
SAPP: signal acquisition and postprocessing
UTC: Coordinated Universal Time
VP: virtual patient
VR: virtual reality

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by A Civit, D Economou, K Glover; comments to author 07.03.20; revised version received 10.04.20; accepted 19.04.20; published 02.09.20.

Please cite as:
Antoniou PE, Arfaras G, Pandria N, Athanasiou A, Ntakakis G, Babatsikos E, Nigdelis V, Bamidis P
Biosensor Real-Time Affective Analytics in Virtual and Mixed Reality Medical Education Serious Games: Cohort Study
JMIR Serious Games 2020;8(3):e17823
URL: http://games.jmir.org/2020/3/e17823/
doi:10.2196/17823
PMID:32876575

Original Paper

Virtual Reality–Based Executive Function Rehabilitation System for Children With Traumatic Brain Injury: Design and Usability Study

Jiabin Shen[1], PhD; Henry Xiang[2], MD, MPH, PhD; John Luna[3], MFA; Alice Grishchenko[3], BFA; Jeremy Patterson[3], BA; Robert V Strouse[3], MFA; Maxwell Roland[3], High School Diploma; Jennifer P Lundine[4,5], PhD; Christine H Koterba[5], PhD; Kimberly Lever[2], MA; Jonathan I Groner[6], MD; Yungui Huang[3], MBA, PhD; En-Ju Deborah Lin[3], PhD

[1]Center for Biobehavioral Health, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, United States
[2]Center for Injury Research and Policy, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, United States
[3]Research Information Solutions and Innovation, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, United States
[4]Department of Speech & Hearing Science, The Ohio State University, Columbus, OH, United States
[5]Inpatient Rehabilitation Program, Nationwide Children's Hospital, Columbus, OH, United States
[6]Trauma Program, Nationwide Children's Hospital, Columbus, OH, United States

Corresponding Author:
En-Ju Deborah Lin, PhD
Research Information Solutions and Innovation
Abigail Wexner Research Institute at Nationwide Children's Hospital
700 Children's Drive
Columbus, OH, 43205
United States
Phone: 1 6143553071
Email: deborah.lin@nationwidechildrens.org

Abstract

Background: Traumatic brain injury (TBI) poses a significant threat to children's health. Cognitive rehabilitation for pediatric TBI has the potential to improve the quality of life following the injury. Virtual reality (VR) can provide enriched cognitive training in a life-like but safe environment. However, existing VR applications for pediatric TBIs have primarily focused on physical rehabilitation.

Objective: This study aims to design and develop an integrative hardware and software VR system to provide rehabilitation of executive functions (EF) for children with TBI, particularly in 3 core EF: inhibitory control, working memory, and cognitive flexibility.

Methods: The VR training system was developed by an interdisciplinary team with expertise in best practices of VR design, developmental psychology, and pediatric TBI rehabilitation. Pilot usability testing of this novel system was conducted among 10 healthy children and 4 children with TBIs.

Results: Our VR-based interactive cognitive training system was developed to provide assistive training on core EF following pediatric TBI. Pilot usability testing showed adequate user satisfaction ratings for both the hardware and software components of the VR system.

Conclusions: This project designed and tested a novel VR-based system for executive function rehabilitation that is specifically adapted to children following TBI.

(JMIR Serious Games 2020;8(3):e16947) doi:10.2196/16947

KEYWORDS

traumatic brain injury; virtual reality; neurological rehabilitation; executive function; cognitive rehabilitation

Introduction

Background

Opportunities to use virtual reality (VR) in health care are increasing as technology advances and accessibility to VR improves. VR provides a novel way to create an engaging virtual world to deliver authentically convincing simulations with measurable tasks in a controlled, low-risk environment. Current health care applications of VR include patient or provider education, pain or anxiety reduction, and therapeutic interventions [1-3].

Leveraging VR for the rehabilitation of cognitive and motor functions for children after a traumatic brain injury (TBI) is an active field of research and development [4-6]. TBI often disrupts the normal function of the brain of a child and is the leading cause of death and acquired disability in children, with an estimated 700,000 pediatric TBI cases annually in the United States [7,8]. Severe TBI in children often causes significant functional losses in memory, communication, and muscle control; long-term therapies are required to recover both cognitively and physically. Most traditional rehabilitation exercises involve repetitive, task-oriented training, which patients with TBI have reported to be boring, leading to poor adherence [9,10]. VR may provide a more engaging and cost-effective alternative or supplement to the traditional rehabilitation program [4-6].

Current therapeutic VR applications for pediatric TBI mostly focus on physical rehabilitation and include games that assist patients in regaining balance, increasing extremity strength, and practicing real-life tasks in low-risk environments [5,11-13]. Our systematic review of the use of VR in pediatric TBI rehabilitation found only 3 studies meeting the review criteria, and all of them focused on the positive effects of VR on physical rehabilitation [6]. However, it is promising to learn that VR's efficacy in cognitive rehabilitation in adults with TBI has turned out positive. For instance, Grealy et al [14] found that exercise-based VR rehabilitation increased adult patient performance on visual and verbal learning tasks as compared with patients in traditional rehabilitation programs. Jacoby et al [15] reported that adults with TBI in VR-based task-specific therapy performed better on executive function tests than patients in therapy without VR. Finally, Caglio et al [16] found that adult patients who used VR for navigational tasks increased their memory capacity.

Objectives

The goal of this study was to design a VR system specifically for executive function (EF) rehabilitation among children with TBI. Few VR systems have been reported in this domain, with even fewer designed specifically to meet the physical, psychological, and medical needs of this vulnerable population. Potential complications of pediatric TBI may dictate the design of an appropriate VR-based cognitive rehabilitation solution for this patient population. For example, most of the current VR systems rely on a head-mounted display (HMD) and require the patient to move his or her head to navigate in the VR environment, which can be unsafe for patients with severe head injuries due to possible skull fractures or scalp sutures. During the process of addressing these safety concerns and developing a VR-based interactive cognitive training system (referred as *the VR system* in the remainder of the paper) to facilitate EF rehabilitation in children with TBI, we identified several useful approaches that may inform the future development of VR applications with similar challenges. This paper presents the specific design elements of the VR system to highlight these practical considerations.

Methods

Design and Development of the Virtual Reality System

Developing a VR solution typically involves many steps and components, including hardware selection and customization, the development and testing environment, defining outcome metrics, end-user interaction design, administrator control design, and data collections and analytics. This section provides a detailed review of the major components of the VR solution.

Hardware Options of the Virtual Reality System

A variety of devices and components can deliver the VR experience and build VR software. In most cases, highly desired features (eg, lightweight and positional tracking) should be considered in concert with the proposed utility and the available resources. Currently, the main categories to consider are smartphone VR headsets, tethered PC-based VR headsets (eg, HTC Vive), or standalone VR headsets (eg, Oculus Quest; Table 1). Smartphone VR headsets deliver the VR experience through a smartphone fitted on a headset that can be as simple as the original Google Cardboard. Although these types of VR applications are more cost-effective and easier to disseminate, resolution, frame rate, and insufficient input mechanisms supporting user interaction limit the delivered VR experience as compared with the higher end PC-based or standalone VR headsets. The tethered VR systems include a headset that is physically or wirelessly connected to a computer. A high-quality headset connected to a powerful gaming PC tends to provide the most immersive VR experience because of the high tracking accuracy and superior graphics quality. Console VR, which is currently limited to Playstation VR, offers features similar to PC-based systems. Unlike PC-based VR, Playstation VR is powered by a Sony Playstation 4 video game console. Positional tracking is performed by a single Playstation camera, and Playstation Move controllers are used for input. The standalone VR headsets (also referred to as all-in-one VR headsets) have built-in processors, sensors, batteries, storage memory, and displays. These systems are wireless and easy to use and typically offer a VR experience of a quality between that provided by the smartphone VR and PC-based VR. Consequently, current technology leaders in VR are focusing more on the design of headsets in this category, and wireless yet powerful VR headsets are likely to dominate soon. Recent offerings such as the Oculus Quest and Vive Focus Plus already adopt the wireless inside-out tracking to provide 6 degrees of freedom (DoF) and have the potential for unlimited movement.

Table 1. Comparison of commercially available virtual reality systems.

Type of headset	Advantages	Disadvantages	Examples
Smartphone VR[a]	• Low cost (ie, viewer using an existing smartphone) • Easier set up • Scalability	• Low-fidelity graphics • Increased work to achieve optimization • Three DoF input • Requires compatible smartphone • Increased motion sickness risk	• Samsung GearVR • Google Daydream[b] • Google Cardboard
PC-based VR	• Immersive presence • Precise 6 DoF[c] tracking (ie, orientation and position) • Room scale • Allows external monitoring • High fidelity	• Heavy • Tethered to PC • Advance setup (outside-in only) • Loss of hand tracking (inside-out only)	• Oculus Rift S • HTC Vive Suite • Valve Index • Windows Mixed Reality
Console VR	• Comfortable headset • Less expensive than gaming PC	• Requires Sony development license and DevNet access • Loses tracking easily • Tethered to console	• Playstation VR
Standalone VR	• Portability • Easier set up • Cheaper to deploy compared with the console or PC-based VR	• Low-fidelity graphics • Increased work to achieve optimization • No or limited movement on 3 DoF hardware	• Oculus Quest • Vive Focus Plus • Oculus Go (3 DoF only) • Lenovo Mirage VR S3 (3 DoF only)

[a]VR: virtual reality.

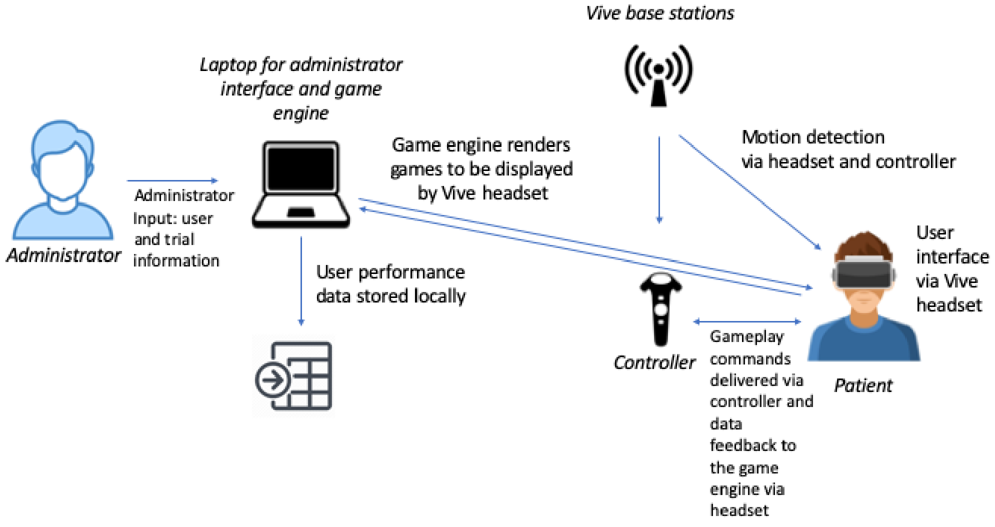
[b]Google recently announced that they will not sell the Daydream viewer and will not include Daydream compatibility in phones going forward. Existing Daydream devices will still have access to the Daydream platform at the time of writing.

[c]DoF: degrees of freedom.

Hardware of the Virtual Reality System

For the purpose of this study, we chose to use a PC-based VR program. The system consists of an HTC Vive VR headset and Vive controller (both by HTC Vive Tech Corporation), an Alienware laptop (Dell Inc), a customized portable station, and 2 infrared projectors with tripods. Figure 1 depicts the interactions of the system components. The HTC Vive system and an Alienware laptop were used as they were tailored for gaming with superior real-time graphics rendering. At the time of its initial development in 2016, the Vive offered the best field of vision and resolution with the most reliable positional tracking. For positional tracking, the Vive base stations emit alternating infrared pulses and laser sweeps at 60 times per second. The photosensors on the headset and controllers use the timing difference between the infrared and lasers to determine the position and orientation with submillimeter precision. A top-of-the-line gaming laptop, such as hardware from Alienware, reduces the likelihood of VR simulation sickness by rendering more complex real-time graphics at a higher frame rate [17], while also allowing for increased portability, as compared with a desktop PC.

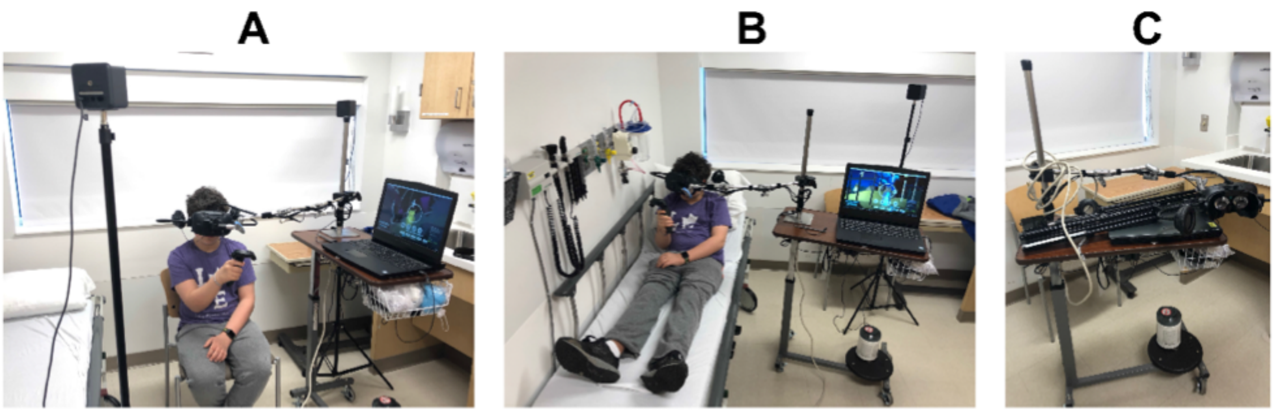
Figure 1. System component diagram showing data flow and user interaction.



A Vive controller input device was used to provide the user with a physical connection to the virtual world. The game displayed the Vive controller as a virtual hand with which the user presses virtual buttons. By seeing their hand movements represented directly in the virtual world, the user had a stronger sense of presence (ie, the sense of being in the virtual environment), thus leading to a more immersive experience [18-20]. However, one important lesson learned was that the large size of the Vive controller relative to the small hands of the younger users led to their use of both hands to handle the controller. This dual hand usage could have led to a dissociation of the one-to-one hand representation for some users.

One unique consideration in designing a VR system for pediatric TBI rehabilitation is minimizing the headset weight upon the child's head when donning the HMD. As the VR headset can weigh as much as 0.83 kg (0.56 kg in the case of the HTC Vive), some secondary effects of a TBI (eg, skull fracture and scalp sutures) might preclude direct head mounting. To circumvent this issue, we custom-mounted the VR headset to an adjustable mechanical arm attached to a cart (Figure 2). Also, a headphone secured on the sides of the headset reduced direct contact and weight on the head. The headset is hence positionally and rotationally fixed and does not weigh down on the head. The mechanical support system was designed to accommodate users in both sitting and reclining positions. This allowed users to experience VR in a chair or in their hospital bed. In this setup, 2 infrared projectors placed on either side of the user served to detect the position of the controller relative to the VR space. A detailed description of the custom-mount setup is provided in Multimedia Appendix 1.

Figure 2. Virtual reality–based interactive cognitive wheeled workstation. Operated with a child sitting in a chair (left). The child operating the program while reclining in bed (middle). Equipment is sanitized after each use and stored as a compact and portable workstation (right).



Software of the Virtual Reality System

Development Environment

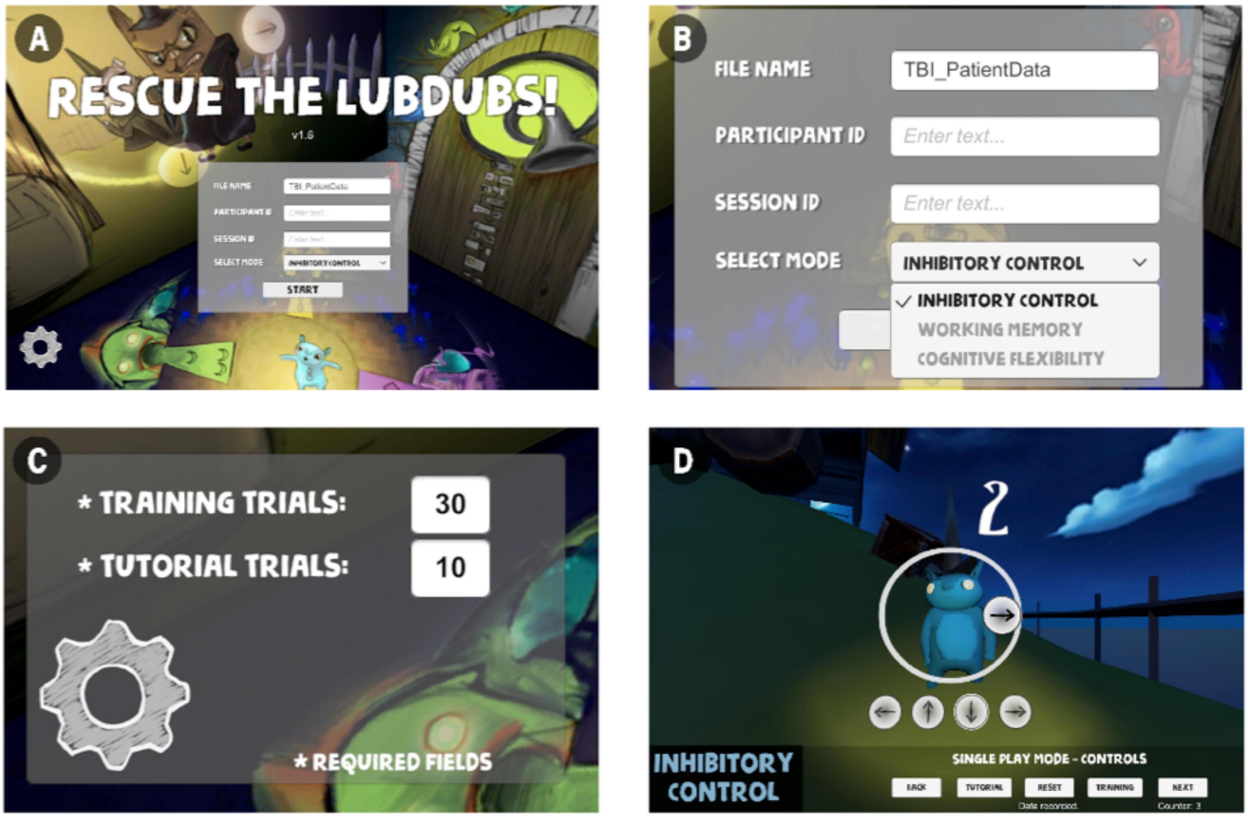
The increasing accessibility of VR affords many software choices for developing VR, such as Unity3D, Unreal Engine, AppGameKit VR, CryEngine, Amazon Lumberyard, and ApertusVR. The choice of software is the developer's preference as they are comparable. For the VR system described in the study, the game contents were developed using the Unity game engine (Unity Technologies). Maya 3D software (Autodesk) was used for 3D modeling and animation, and Photoshop (Adobe) was used to create 2-dimensional assets.

Interface for the Researcher or Therapist

As cognitive rehabilitation is a long-term repetitive process, designing an interface that allows the therapist to enter patient and session information to track progress is an important consideration. Enabling specific controls over the game, such as stop and restart, and options to choose different modules is also advantageous. Likewise, the ability to observe patients as they play the game may give the therapist insights into their progress and deliver timely feedback or support.

One of the advantages of a PC-tethered system (eg, Vive) is that it easily accommodates a separate interface for the therapist without needing to connect separate devices over a server. Through the main interface for the VR system (Figure 3), the therapist can enter user information and session IDs, select the training module, customize training by setting the number of trials the user will play, and monitor VR training progress. The therapist can control the progress of the game on the interface with 5 buttons—*Back* (go back to the home screen), *Tutorial* (a short version of the trial, data not collected), *Reset* (interrupts the trial and resets the trial to its initialized state), *Training* (full-version trial with data collection), and *Next* (go to the next game). The therapist can directly instruct the user in the tutorial module by drawing in the virtual space using the trackpad on the laptop or a second controller.

Figure 3. User-friendly interface for therapist. (A) Enter user ID and session ID. (B) Choose training module to run. (C) Customize training by changing number of trials. (D) Control, track, and instruct using laptop trackpad or use a second controller.



Virtual Reality Game Design for Executive Function Training

Overview

Game design is a crucial element in creating an effective VR cognitive rehabilitation system. Cognitive exercises with adjustable levels of difficulty are required to meet the varied needs of TBI users. More in-depth reviews of various game designs for cognitive training are available for the interested reader [21-23]. In this section, we present the development of the VR system to highlight some design considerations.

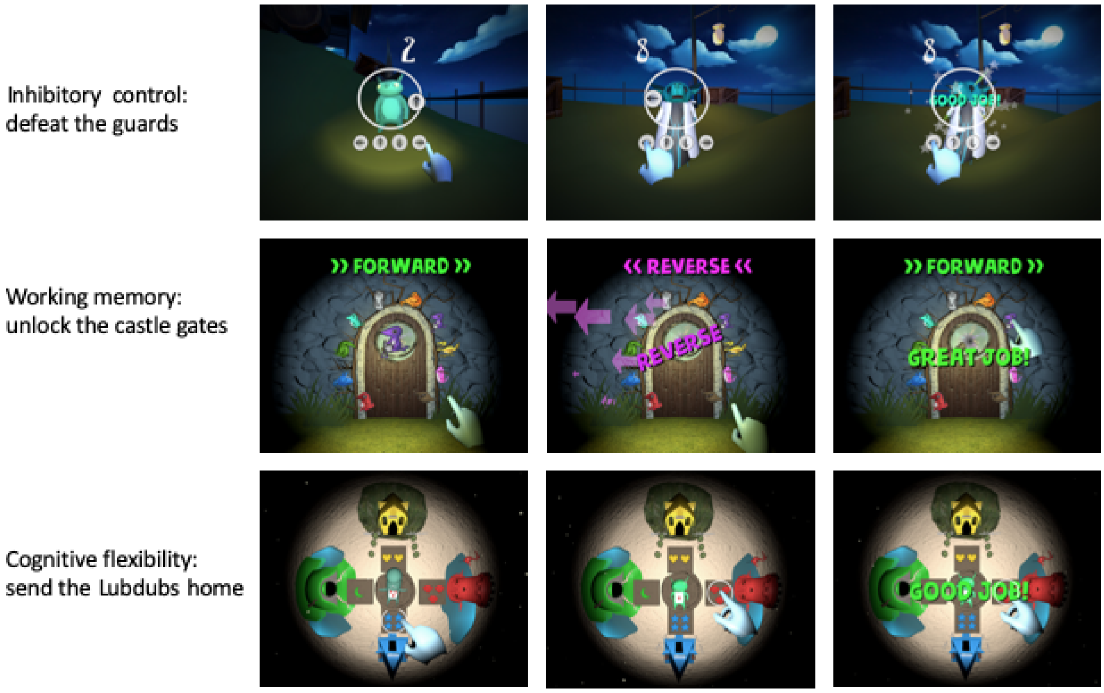
Development of our VR system was focused on executive function rehabilitation because pediatric TBIs, especially moderate to severe cases, often result in executive dysfunction due to the vulnerability of the frontal lobes. One theoretical rationale underlying the mechanism for repeated VR-based tasks to have a potential training effect on the EF of children with TBI beyond the trained tasks is that EF are mostly regarded as domain-general skills, as opposed to domain-specific skills, in both a healthy population [24,25] and patients with TBI [26,27]. Following this rationale, the VR system includes 3 VR games for training 3 core EF: game 1 for *inhibitory control* (the ability to override a strong internal predisposition or external lure and do what is more appropriate or needed), game 2 for *working memory* (the ability to hold and process information in mind as needed), and game 3 for *cognitive flexibility* (the ability to adjust to changing environmental demands and think from different perspectives).

Story Narrative

A useful and increasingly popular approach in digital game design is the use of a story narrative [28-30]. A story narrative creates a more meaningful and immersive experience for the user. A narrative background helps the user to feel more involved and thus enhances engagement. In addition, background narrative is a mechanism that conveys the perceived conceptual depth of the virtual environment to the user before active engagement begins, thus reinforcing the user's sense of realism. The story narrative, presented at the start of the game, is best delivered through voice narration rather than text, to improve access for a younger target audience. As illustrated in Figure 4, the story narrative for the VR system is a mission to *Rescue the Lubdubs* and presented by the research staff as follows:

> *Lubdubs are magical creatures that live in a different world. They have been captured, and your job is to return them safely to their homes. You will play three mini-games. Our goal is to get through all the guards of the castle by completing each of the three games and rescuing the lubdubs within the castle.*

Figure 4. Virtual reality–based interactive cognitive training modules. (Top panel) inhibitory control, (middle panel) working memory, and (bottom panel) cognitive flexibility.



Game 1: Inhibitory Control

This game design was based on a classic psychological task for inhibitory control, the Spatial Stroop Task [31], which is used to train and assess inhibitory control. The Spatial Stroop Task examines the ability of the user to respond correctly to the interference between the stimulus location with the location information in the stimuli. In this game, the user is *battling* different characters. An arrow appears randomly on 1 of the 4 positions in relation to the character the user is *battling* (ie, above, below, to the left, or to the right). The direction of the arrow may or may not be the same as its *position* relative to the character. The user is required to tap on the arrow below that matches the arrow that appears on the circle in the middle of the screen. Inhibitory control is the ability to override a strong internal predisposition or external cues and do what is more appropriate. Thus, in this game, the user needs to ignore the *positional* cue (eg, right of the character) and respond to the actual direction of the arrow (eg, up arrow, as depicted in Figure 4 top panel). To maintain interest and engagement, the user will move to a different character after a few rounds, as if progressing through a series of battles. The data collected for this module include time taken to respond, arrow position and direction, user response (arrow direction), and correctness of the selection.

Game 2: Working Memory

This game is a sequence recall game for training working memory that is adapted from the Visual Working Memory Task [32]. This game consists of a locked door with different characters around the door (Figure 4 middle panel). To *unlock* the door, the user needs to remember the order of characters displayed on the center of the door. The game starts with a sequence of 2 characters and adjusts difficulty level based on the responses of the user. Every time the user gets 2 consecutive sequences correct, the sequences increase in length by 1 character.

Conversely, if 2 consecutive sequences are incorrect, the sequences reduce in length by 1. Also, the trial will ask the user to recall the displayed sequence either in forward or reverse order. This reverse task is an essential component of working memory training. By definition, working memory is a system for temporary information storage for the execution of more complex cognitive tasks such as reasoning and information manipulation. The reverse task is designed to require an additional level of information processing.

To sustain interest, the user progresses through 5 different doors, with different characters, performing one-fifth of the total trials at each door. The sequence length carries over from the previous trial (ie, does not reset at each door) to ensure the trial will be sufficiently challenging to achieve the training purpose. The data collected for the module include time to submit the sequence, length of the presented sequence, length of the user sequence, and accuracy of the sequence.

Game 3: Cognitive Flexibility

This game on cognitive flexibility is adapted from the Wisconsin Card Sorting Task [33,34]. In this task-switching paradigm, the user has to figure out which rule underlies a task (eg, match by color, shape, or number) and determine when the rule changes. In the VR system, to send the Lubdubs back to their homes, the user has to figure out what sorting method is being used to match the symbol on the Lubdub's stomach to one of the four symbols in front of the houses (as shown in Figure 4 bottom panel). Every choice will result in a correct or incorrect feedback from the program to the user, and the user will need to determine the current rule based on the response. The rule is set to change every 7 trials, although this is undisclosed to the user. The data collected for this game include the time taken to respond, presented icon, current matching rule, user choice, and the accuracy of the answer.

Usability Testing of the Virtual Reality System

We conducted feasibility exercises of the VR system (the same version as used for patient population) on a convenience sample of healthy volunteers before testing on users with TBI. As the data collected were intended to be used as feedback data for the developers to improve the design of the VR systems rather than for human research purposes, the institutional review board (IRB) at Nationwide Children’s Hospital agreed that IRB review was not required. This feasibility exercise among healthy volunteers allowed us to better understand and potentially reduce the likelihood of negative effects on the pediatric users with TBI who may have been more sensitive to particular sensory effects (eg, simulation sickness and discomfort due to sound and light effects). In total, 10 healthy children aged between 7 and 17 years (mean age 14.30 years, SD 3.56 years) recruited from local communities in a midwestern urban city participated in this feasibility exercise. After playing for 10 min (which was sufficient time to try out all 3 games, although they were not required to complete all trials of each game as per the goal of this feasibility exercise), the users filled out the Simulation Sickness Questionnaire (SSQ) [35]; the Borg Perceived Physical Exertion Scale [36]; and a brief custom-made VR experience survey with questions on pleasure, motivation, and realism (Multimedia Appendix 2). For each of the questions, the users were asked to rate their responses from 0 (not at all) to 100 (very much). We also examined how patients with TBI completed the tasks, the time taken to complete the tasks, and the accuracy of responses. The protocol for collecting data from patients with TBI was reviewed and approved by the IRB, and informed consent or assent was obtained from the patients and their legal guardians before participating in study activities. A total of 4 pediatric patients with moderate to severe TBI as determined by the Glasgow Coma Scale (mean 5.00, SD 3.05) were recruited from an inpatient rehabilitation unit in a level I trauma center in a midwestern urban city. The patients were aged between 7 and 17 years (mean 11.75 years, SD 2.60 years), and all of them completed up to 50 trials for each of the 3 VR tasks and the same set of usability surveys as the healthy users.

Data collected by the VR system (Table 2) are stored locally on a password-protected laptop in a csv file without Protected Health Information connected to any of the data. These raw data were used to derive the response time to complete each trial and the percentage of correct responses.

Table 2. Metrics data automatically collected by virtual reality–based interactive cognitive training.

Task and metrics[a]	Possible values
Inhibitory control	
Location of arrow	(up or down or left or right)
Direction of arrow	(up or down or left or right)
Condition	(consistent or inconsistent)
Response	(up or down or left or right)
Response time	(minutes:seconds:milliseconds)
Correct?	(yes or no)
Working memory	
Order	(forward or backward)
Number of items	(provide range)
Level	(provide range)
Response time	(minutes:seconds:milliseconds)
Correct?	(yes or no)
Cognitive flexibility	
Rule applied	(amount or shape)
Amount or color or shape	(amount or color or shape)
Response	(amount or color or shape)
Response time	(minutes:seconds:milliseconds)
Correct?	(yes or no)

[a]For each training session, the study ID and the session number are entered by the researcher and are stored along with the metrics collected for the 3 training modules.

Results

Testing in Healthy Children

As shown in Table 3, healthy children reported a high level of fun and enjoyability. In the context of therapy, participants reported that they would highly desire that the VR system be available at hospitals and would be highly motivated to attend postdischarge therapies if they were ever to have a TBI. The participants reported low levels of simulation sickness on the SSQ and felt very light exertion due to playing the VR games.

Table 3. Usability findings in healthy children and children with traumatic brain injury.

Measure	Healthy children (n=10), mean (SD)	Children with TBI[a] (n=4), mean (SD)
How realistic did you feel about the virtual reality environment? (0-100)	58.8 (35.6)	37.5 (47.9)
How much fun did you feel about the virtual reality games you just played? (0-100)	72.0 (23.9)	73.5 (31.5)
How did you like the virtual reality games you just played? (0-100)	72.0 (25.8)	67.5 (39.5)
Do you want to play them again in future? (0-100)	72.5 (26.8)	56.3 (43.9)
Do you want to have such virtual reality games in your future therapies while you are in hospital? (0-100)	85.0 (22.1)	50.3 (52.5)
Would you be more motivated to attend your therapy sessions after discharge if we include such games? (0-100)	87.0 (19.9)	41.5 (49.9)
Simulator sickness (0-48)	1.60 (1.07)	2.50 (2.08)
Physical exertion (6-20)	8.25 (1.63)	9.9 (2.3)

[a]TBI: traumatic brain injury.

As seen in Table 4, the healthy children were generally able to complete the trials in a reasonable amount of time (averaging less than 4 seconds per trial). The percentage of correct responses indicated that healthy children were able to complete inhibitory controls with high accuracy, whereas the other 2 tasks, working memory and cognitive flexibility, proved to be more challenging.

Table 4. Virtual reality–based interactive cognitive training performance of healthy children and pediatric patients with traumatic brain injury.

Task	Number of trials completed, range		Average time per trial, seconds, mean (SD)		Proportion of correct responses[a] (%), mean (SD)	
	Healthy children[b] (n=8)	Children with TBI[c] (n=4)	Healthy children[b] (n=8)	Children with TBI (n=4)	Healthy children[b] (n=8)	Children with TBI (n=4)
Inhibitory control	11-50	50[d]	3.2 (5.5)	4.9 5.6)	98 (3.2)	95 (7.6)
Working memory	10-34	30-50	3.9 (1.1)	11.8 (9.2)[e]	74 (5.5)	50 (17.1)
Cognitive flexibility	32-50	50[d]	2.7 (1.6)	6.2 (4.7)	71 (11.9)	59 (9.2)

[a]Percentage of correct responses is calculated as the number of trials with correct responses over the total number of trials completed for the task by the subject.

[b]Data were not captured properly for 2 out of the 10 healthy volunteers, and thus, they were omitted.

[c]TBI: traumatic brain injury.

[d]Everyone completed 50 trials.

[e]Two users completed only 30 trials of the working memory task due to technical issues of the system, not due to their inability to complete.

Testing in Children With Traumatic Brain Injury

Our preliminary testing in 4 children with TBI indicates that children with TBI can complete all 3 games, although requiring longer time for each trial (Table 4). On the basis of the trial response time captured in the app, the cumulative completion time for 50 trials ranged from 63.0 to 670.1 seconds for the inhibitory control task, 347.9 to 1283.5 seconds for the working memory task, and 92.2 to 537.7 seconds for the cognitive flexibility task. However, the loading and feedback time between trials was not captured, so the actual time taken to participate in the training was longer. There were significant subject variations in task performance, highlighting substantial variability in the cognitive ability of the patients (also reflecting the small sample size of this pilot testing), which is to be expected due to the wide range of effects of TBI on different individuals. The working memory task appears to be the most challenging among the 3 tasks, with low accuracy scores and long response times.

In the pilot study, the number of trials for each task was set at 50 by the researchers. This number was based on the estimated time taken to complete each trial for children within reasonable tolerance levels. The intention was to set the number of trials at a level achievable within a reasonable amount of time (15 min) to prevent overfatigue and disengagement. Concurrently, it should also be long enough to achieve the *training* purpose.

With the variation in the cognitive ability in children with TBI, it would be useful to include features for the therapists to adjust the number of trials and/or difficulty of training for an individual patient so that training is sufficiently challenging yet achievable (Figure 3) [10]. The optimal training remains to be evaluated in clinical studies and is outside the scope of this paper.

Children with TBI also completed the same questionnaires on usability and the SSQ. As with healthy subjects, they reported low levels of simulation sickness on the SSQ and felt very light exertion due to playing the VR games (Table 3). Interestingly, although children with TBI also reported a high level of fun engagement with the VR games, their attitude toward future use in therapy was highly variable. This may be, in part, due to the rehabilitative nature of the games being more challenging for these children, as indicated by longer trial completion time and lower percentage response accuracy (Table 4), in addition to the small sample size. Notably, some children were anecdotally comparing the games with commercial video games and not with the current rehabilitative programs. Future studies should assess children's preference for VR rehabilitation against standard rehabilitation programs.

For both healthy and pediatric TBI groups, along with the structured quantitative assessment, this study used unstructured verbal feedback by the children and observation during the testing to inform further adjustment to the system. For example, spacing for selection choices in the *working memory* game was observed to be too narrow for some children and was adjusted to reduce frustration in aiming for their intended selection. It was observed that unintentional selection was made when the controller hovers over the selection choices. To address this, the control mechanism changed to mimic the movement of the hand actually needed to represent pressing a physical button. This was later replaced again to a laser pointer style *click* to reduce the complexity. It was also observed that children with smaller hands tend to hold the controller with both hands and thus have a different way of interacting with the controller. This made it hard for these children to reach the trigger button on the Vive controller, so we duplicated all trigger button functions to the trackpad button.

Discussion

Design Considerations for Virtual Reality−Based Cognitive Rehabilitation in Pediatric Patients With TBI

The development of the VR system offered a unique cross-disciplinary perspective, incorporating expertise from professionals working in developmental psychology, digital health, and pediatric rehabilitation. A user-centered design philosophy was implemented to create custom-developed hardware and software systems that indicated a high degree of usability for pediatric patients following a TBI. Textbox 1 summarizes a series of practical considerations specific to designing VR systems for pediatric TBI cognitive rehabilitation that we encountered and solved in developing this system.

Textbox 1. Design considerations for virtual reality applications used for traumatic brain injury rehabilitation.

Headset burden on the injured head

- Mounted virtual reality (VR) headset on mechanical arm frees pediatric patients with traumatic brain injury (TBI) from physical burden

VR simulation sickness

- Mounted headset allows the virtual world to remain relatively static to minimize simulation sickness for pediatric patients with TBI as compared with the traditional 360° VR environment
- Dimming the peripheral view and "lighting up" the center of the gaming screen reduces the vision field
- A high-quality tracking system and high frame rate of the graphics help reduce motion sickness

Engagement or replayability

- Games designed with varying levels of difficulty help engage patients with different baseline skills and gaming experience
- Built-in procedurally generated design elements increase engagement and replayability by introducing near-infinite variety to each gameplay session. Example: various components of a character (eg, head, face, body, and color) can be randomly assembled to generate many different characters quickly

User experience

- Spacing between selection options can affect the user experience. This ensures users can easily select an answer

Rehabilitation factors

- Evidence-based rehabilitation theories in cognitive and developmental psychology and other fields guided the game designs
- Data collection and analysis were used, which are critical to track progress in long-term rehabilitation programs
- Adjustable levels of training matched with individual needs and progress to allow for improving functions in rehabilitation settings

Privacy and data security

- Implemented best practices in personal health data storage and security based on the data collected to ensure the highest level of privacy and security

Limitations and Future Directions

This study has several limitations that we hope to overcome in future studies. First, this study used a convenience sample of healthy children and pediatric patients with TBI, which may not be representative of the target population in age and gender distributions. Second, this study was intended to collect usability data rather than outcome evaluation; therefore, we did not test out the feature of changing number of trials as a way to adjust the level of training difficulty. Third, for children with TBI, the initial intention was to have the subject complete the trials in a single session. However, in practice, due to the preferences of patients and time constraints from their other appointments, it was more feasible to complete all trials in multiple sessions. However, this information was not collected or analyzed in this preliminary study. Future studies of formal efficacy evaluation should consider adding this information in their analysis. Finally, future research should consider design features that are specific for patients in different age groups and with gaming experiences (eg, developing different story narratives) or that have varying number of training components for different subgroups of the pediatric population to increase both the level of engagement and potentially the training efficacy of the program. A pilot randomized clinical trial is currently underway to evaluate the preliminary efficacy of the VR system.

Conclusions

VR offers an exciting approach for pediatric TBI cognitive rehabilitation. This tutorial describes the challenges faced, solutions to these problems, and lessons learned through the development of the VR system. With rapid advances in VR technology and accessibility, we believe there is significant potential to expand the current program for future telemedicine or in-home applications. However, more studies are needed to refine the design of these technologies and evaluate their feasibility and efficacy.

Acknowledgments

The authors would like to thank Colleen Vincent for providing valuable insights into the initial conceptualization of the training content, Madeleine Schroeder for assisting with the literature review, Megan Armstrong for assisting with patient recruitment and data collection, and Dr H Gerry Taylor and Dr Melody Davis for assisting with manuscript edits. This study was supported by Nationwide Children's Hospital Intramural Fund Program, the Clinical and Translational Science Awards (grant number: UL1TR002733). Part of the efforts by JS in analyzing data and writing this paper was also supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development of the National Institutes of Health under award number

K99HD093814. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed hardware configurations on the VR mount setup.
[DOCX File , 15 KB - games_v8i3e16947_app1.docx]

Multimedia Appendix 2
Brief VR Experience Survey.
[PDF File (Adobe PDF File), 58 KB - games_v8i3e16947_app2.pdf]

References

1. Dumay AC. Medicine in virtual environments. Technol Health Care 1995 Oct;3(2):75-89. [Medline: 8574765]
2. Moline J. Virtual reality for health care: a survey. Stud Health Technol Inform 1997;44:3-34. [Medline: 10175341]
3. Riva G. Applications of virtual environments in medicine. Methods Inf Med 2003;42(5):524-534. [Medline: 14654887]
4. Aida J, Chau B, Dunn J. Immersive virtual reality in traumatic brain injury rehabilitation: a literature review. NeuroRehabilitation 2018;42(4):441-448. [doi: 10.3233/NRE-172361] [Medline: 29660958]
5. Pietrzak E, Pullman S, McGuire A. Using virtual reality and videogames for traumatic brain injury rehabilitation: a structured literature review. Games Health J 2014 Aug;3(4):202-214. [doi: 10.1089/g4h.2014.0013] [Medline: 26192369]
6. Shen J, Johnson S, Chen C, Xiang H. Virtual reality for pediatric traumatic brain injury rehabilitation: a systematic review. Am J Lifestyle Med 2020;14(1):6-15 [FREE Full text] [doi: 10.1177/1559827618756588] [Medline: 31903073]
7. Keenan HT, Bratton SL. Epidemiology and outcomes of pediatric traumatic brain injury. Dev Neurosci 2006;28(4-5):256-263. [doi: 10.1159/000094152] [Medline: 16943649]
8. Report to Congress on Traumatic Brain Injury in the United States: Epidemiology and Rehabilitation. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/traumaticbraininjury/pubs/congress_epi_rehab.html [accessed 2020-02-04]
9. Burdea GC. Virtual rehabilitation--benefits and challenges. Methods Inf Med 2003;42(5):519-523. [Medline: 14654886]
10. Shapi'i A, Zin NA, Elaklouk AM. A game system for cognitive rehabilitation. Biomed Res Int 2015;2015:493562 [FREE Full text] [doi: 10.1155/2015/493562] [Medline: 25815320]
11. Biffi E, Beretta E, Cesareo A, Maghini C, Turconi AC, Reni G, et al. An immersive virtual reality platform to enhance walking ability of children with acquired brain injuries. Methods Inf Med 2017 Mar 23;56(2):119-126. [doi: 10.3414/ME16-02-0020] [Medline: 28116417]
12. da Silva CM, Badia SB, Duarte E, Verschure PF. Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: a randomized controlled pilot study in the acute phase of stroke using the rehabilitation gaming system. Restor Neurol Neurosci 2011;29(5):287-298. [doi: 10.3233/RNN-2011-0599] [Medline: 21697589]
13. Ustinova KI, Perkins J, Leonard WA, Hausbeck CJ. Virtual reality game-based therapy for treatment of postural and co-ordination abnormalities secondary to TBI: a pilot study. Brain Inj 2014;28(4):486-495. [doi: 10.3109/02699052.2014.888593] [Medline: 24702281]
14. Grealy MA, Johnson DA, Rushton SK. Improving cognitive function after brain injury: the use of exercise and virtual reality. Arch Phys Med Rehabil 1999 Jun;80(6):661-667. [doi: 10.1016/s0003-9993(99)90169-7] [Medline: 10378492]
15. Jacoby M, Averbuch S, Sacher Y, Katz N, Weiss PL, Kizony R. Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: a pilot study. IEEE Trans Neural Syst Rehabil Eng 2013 Mar;21(2):182-190. [doi: 10.1109/TNSRE.2012.2235184] [Medline: 23292820]
16. Caglio M, Latini-Corazzini L, D'Agata F, Cauda F, Sacco K, Monteverdi S, et al. Virtual navigation for memory rehabilitation in a traumatic brain injured patient. Neurocase 2012;18(2):123-131. [doi: 10.1080/13554794.2011.568499] [Medline: 22352998]
17. DiZio P, Lackner JR. Circumventing Side Effects of Immersive Virtual Environments. Semantic Scholar. 1997. URL: https://www.semanticscholar.org/paper/Circumventing-Side-Effects-of-Immersive-Virtual-DiZio-Lackner/8706f29f070ba9f5ba98b1ff4c81e2404a7b634e [accessed 2020-06-25]
18. Kilteni K, Groten R, Slater M. The sense of embodiment in virtual reality. Presence-Teleop Virt 2012 Nov;21(4):373-387 [FREE Full text] [doi: 10.1162/pres_a_00124]
19. Peperkorn HM, Diemer JE, Alpers GW, Mühlberger A. Representation of patients' hand modulates fear reactions of patients with spider phobia in virtual reality. Front Psychol 2016;7:268 [FREE Full text] [doi: 10.3389/fpsyg.2016.00268] [Medline: 26973566]

20. Schultze U. Embodiment and presence in virtual worlds: a review. J Inf Technol 2010 Dec 1;25(4):434-449 [FREE Full text] [doi: 10.1057/jit.2010.25]
21. Zaki A, Wook T, Ahmad K. Therapeutic serious game design guidelines for stimulating cognitive abilities of children with speech and language delay. J Inf Commun Technol 2017;16(2):284-312 [FREE Full text]
22. Baniqued PL, Kranz MB, Voss MW, Lee H, Cosman JD, Severson J, et al. Cognitive training with casual video games: points to consider. Front Psychol 2014 Jan 7;4:1010 [FREE Full text] [doi: 10.3389/fpsyg.2013.01010] [Medline: 24432009]
23. Neugnot M, Gagner C, Beauchamp MH. The use of games in paediatric cognitive intervention: a systematic review. Int J Phys Med Rehabil 2015 Jun 25;3(4):286 [FREE Full text] [doi: 10.4172/2329-9096.1000286]
24. Kroesbergen E, van Luit J, van Lieshout E, van Loosbroek E, van de Rijt B. Individual differences in early numeracy. J Psychoeduc Assess 2009 Apr 7;27(3):226-236. [doi: 10.1177/0734282908330586]
25. Zelazo P, Cunningham W. Executive function: mechanisms underlying emotion regulation. In: Gross J, editor. Handbook of Emotion Regulation. New York, USA: Guilford; 2007:135-158.
26. Dennis M, Agostino A, Roncadin C, Levin H. Theory of mind depends on domain-general executive functions of working memory and cognitive inhibition in children with traumatic brain injury. J Clin Exp Neuropsychol 2009 Oct;31(7):835-847. [doi: 10.1080/13803390802572419] [Medline: 19221924]
27. Gioia GA, Isquith PK. Ecological assessment of executive function in traumatic brain injury. Dev Neuropsychol 2004;25(1-2):135-158. [doi: 10.1080/87565641.2004.9651925] [Medline: 14984332]
28. Abbas NB, Ryan M. Narrative as virtual reality: immersion and interactivity in literature and electronic media. J Engl Stud 2004;34:277 [FREE Full text] [doi: 10.2307/3509518]
29. Lu AS, Thompson D, Baranowski J, Buday R, Baranowski T. Story immersion in a health videogame for childhood obesity prevention. Games Health J 2012 Feb 15;1(1):37-44 [FREE Full text] [doi: 10.1089/g4h.2011.0011] [Medline: 24066276]
30. Dickey MD. Game design narrative for learning: appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. Educ Tech Res Dev 2006 Jun;54(3):245-263 [FREE Full text] [doi: 10.1007/s11423-006-8806-y]
31. Lu CH, Proctor RW. The influence of irrelevant location information on performance: a review of the Simon and spatial stroop effects. Psychon Bull Rev 1995 Jun;2(2):174-207. [doi: 10.3758/BF03210959] [Medline: 24203654]
32. Baddeley A. Working memory: looking back and looking forward. Nat Rev Neurosci 2003 Oct;4(10):829-839. [doi: 10.1038/nrn1201] [Medline: 14523382]
33. Ord JS, Greve KW, Bianchini KJ, Aguerrevere LE. Executive dysfunction in traumatic brain injury: the effects of injury severity and effort on the Wisconsin card sorting test. J Clin Exp Neuropsychol 2010 Feb;32(2):132-140. [doi: 10.1080/13803390902858874] [Medline: 19484646]
34. Puente A. Wisconsin card sorting test. Test critiques 1985;4:677-682. [doi: 10.1007/springerreference_184562]
35. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
36. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, IL: Human Kinetics; 1998.

Abbreviations

EF: executive functions
HMD: head-mounted display
IRB: institutional review board
SSQ: simulation sickness questionnaire
TBI: traumatic brain injury
VR: virtual reality

Edited by T Rashid Soron; submitted 06.11.19; peer-reviewed by J Salisbury, H Sandhu; comments to author 16.12.19; revised version received 14.04.20; accepted 24.04.20; published 25.08.20.

Please cite as:
Shen J, Xiang H, Luna J, Grishchenko A, Patterson J, Strouse RV, Roland M, Lundine JP, Koterba CH, Lever K, Groner JI, Huang Y, Lin EJD
Virtual Reality–Based Executive Function Rehabilitation System for Children With Traumatic Brain Injury: Design and Usability Study
JMIR Serious Games 2020;8(3):e16947
URL: http://games.jmir.org/2020/3/e16947/
doi:10.2196/16947
PMID:32447275

Original Paper

Relationship Between Children's Enjoyment, User Experience Satisfaction, and Learning in a Serious Video Game for Nutrition Education: Empirical Pilot Study

Ismael Edrein Espinosa-Curiel[1], PhD; Edgar Efrén Pozas-Bogarin[1], BSc; Juan Martínez-Miranda[2], PhD; Humberto Pérez-Espinosa[2], PhD

[1]Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad de Transferencia Tecnológica Tepic, Tepic, Nayarit, Mexico

[2]Consejo Nacional de Ciencia y Tecnología, Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad de Transferencia Tecnológica Tepic, Tepic, Nayarit, Mexico

Corresponding Author:
Ismael Edrein Espinosa-Curiel, PhD
Centro de Investigación Científica y de Educación Superior de Ensenada
Unidad de Transferencia Tecnológica Tepic
Andador 10, entre calles 3 y 4, Ciudad del conocimiento
Tepic, Nayarit, 63173
Mexico
Phone: 52 1295930 ext 28607
Email: ecuriel@cicese.edu.mx

Abstract

Background: The design and use of serious video games for children have increased in recent years. To maximize the effects of these games, it is essential to understand the children's experiences through playing. Previous studies identified that enjoyment and user experience satisfaction of the players are principal factors that can influence the success of serious video games and the learning of their players. However, research about the relationship between enjoyment and user experience satisfaction with learning in children 8 to 10 years old is sparse.

Objective: We examined the relationship of enjoyment and user experience satisfaction with the learning of children aged 8 to 10 years while playing a serious video game for health, FoodRateMaster. This serious video game teaches children about the characteristics of healthy and unhealthy foods and how to identify them in their environment.

Methods: Children aged 8 to 10 years were recruited from a primary school in Mexico. Participants completed 12 individual gaming sessions with FoodRateMaster in 6 weeks. A food knowledge questionnaire was administered before and after game play to assess the players' food knowledge. In addition, after the gaming sessions, the children's enjoyment and user experience satisfaction were evaluated using the EGameFlow questionnaire and the Game User Experience Satisfaction Scale (GUESS) questionnaire.

Results: We found significant positive associations for children's (n=60) posttest knowledge with enjoyment (r_{58}=0.36, *P*=.005) and user experience satisfaction (r_{58}=0.27, *P*=.04). The children's posttest knowledge scores were also positively correlated with challenge (r_{58}=0.38, *P*=.003), knowledge improvement (r_{58}=0.38, *P*=.003), and goal clarity (r_{58}=0.29, *P*=.02) EGameFlow subscales and with narrative (r_{58}=0.35, *P*=.006), creative freedom (r_{58}=0.26, *P*=.04), and visual esthetics (r_{58}=0.32, *P*=.01) GUESS subscales. Regression analysis indicated that the EGameFlow ($F_{7,52}$=2.74, *P*=.02, R^2=0.27) and the GUESS ($F_{8,51}$=2.20, *P*=.04, R^2=0.26) ratings significantly predicted the children's posttest knowledge scores. EGameFlow challenge (β=0.40, t_{52}=2.17, *P*=.04) and knowledge improvement (β=0.29, t_{52}=2.06, *P*=.04) subscales significantly contributed to predicting children's learning. None of the GUESS subscales significantly contributed to predicting children's learning.

Conclusions: The findings of this study suggest that both enjoyment and user experience satisfaction for children aged 8 to 10 years were positively correlated with their learning and that were significant predictors of it. Challenge, knowledge improvement, narrative, creative freedom, and visual esthetics subscales correlated positively with children's learning. In addition, challenge and knowledge improvement contributed to predicting their learning. These results are relevant to consider during the design stages of serious games developed for young children's learning purposes.

(JMIR Serious Games 2020;8(3):e21813) doi:10.2196/21813

KEYWORDS

Serious video games; children; enjoyment; user experience; satisfaction; game-based learning; nutrition; serious game; pilot study

Introduction

Serious Video Games for Children

Serious video games are interactive digital programs or apps that have a goal beyond entertainment or fun. This type of video game is designed for promoting and encouraging a broad range of purposes such as education, training, skill development, knowledge acquisition, or even behavior and attitude changes [1,2]. The primary rationale of using video games for serious purposes is to take advantage of their ability to motivate, provide enjoyment, and engage players [1]. In addition, their versatility allows them to be used as tools in a broad spectrum of domains such as the military, government, educational, corporate, enterprise, tourism, and health [3]. Most serious games have reported positive outcomes, and in general, they are more effective than conventional instructional methods [4-7].

The development and use of serious video games designed for children have increased over the past decade [8-11]. Researchers have encouraged their use in children, since children (regardless of age, sex, and background) are interested in video games [9], and because video games allow them to practice different skills in a safe environment. In addition, video games can be used as an alternative to influence children who are having difficulties learning with traditional education methods [1]. There are currently serious video games for children for a broad set of applications, such as STEM education, obesity prevention, nutrition education, chronic diseases, psychiatry, learning disabilities, and cognitive development [9-14].

While there is a growing body of literature on the potential of serious video games in children [9-15], researchers have tended to ignore what children experience during playing sessions, and how these experiences influence the success, effectiveness, and outcomes of serious games. Having a better understanding of their experiences can help in designing more effective and engaging serious video games for children [16-20].

Enjoyment and User Experience Satisfaction in Serious Games

Two of the best-known factors that influence the acceptance, use, and success of any video game are the players' enjoyment and the players' experience satisfaction [21-26]. If players do not enjoy the game or have a satisfying gaming experience, they will not play the game [27]. However, there is a lack of consensus in the video game field regarding the definitions of enjoyment and user experience satisfaction [21,24]. Moreover, as terms, they are often used interchangeably with a host of other terms, some of which overlap conceptually (eg, engagement, flow, fun, playability, and immersion) [21,24].

According to Crutzen et al [21] *enjoyment* refers to "the action or state of deriving gratification from a video game." Enjoyment is a multidimensional concept that embraces dimensions such as concentration, goal clarity, feedback, challenge, autonomy, immersion, social interaction, knowledge improvement, competence, narrative transportation, and relevance [21,27,28]. Meanwhile, Phan et al [24] defined game user *experience satisfaction* as "the degree to which the players feel gratified with his or her experience while playing a video game." Game user experience satisfaction involves different dimensions such as usability (or playability), narratives, creative freedom, audio and visual esthetics, and personal gratification [23,24,26].

The dimensions of enjoyment and player experience satisfaction are important factors for the success and effectiveness of serious games [17,21,29] since boring or poorly designed serious video games run the risk of disappointing and eventually alienating their target audience [30]. Previous studies have identified that factors such as technological capacity, esthetic presentation, game design elements, and fun factors (eg, characters, dialogues, and humor) influence the acceptability of serious games [17]. In addition, factors such as game goals, freedom, narrative, interaction, challenge, sensation, feedback, level of fun, usability, and mystery are critical for the effectiveness of serious games [29,31]. These factors also can influence the mood [19], engagement [21,32], satisfaction [33-35], and performance [34,36] of players. Serious video games can provide excellent levels of usability [35,37] and enjoyment [38] in comparison with entertainment video games [17,39].

Enjoyment, User Experience Satisfaction, and Learning in Serious Games

The studies [20,40-44] that describe the relationship between enjoyment and user experience satisfaction with the learning of players in serious games show contradictory conclusions. While some studies have stated that enjoyment could be perceived as an outcome opposite to learning in serious games [40,41], other studies noted that enjoyment and user experience satisfaction could positively impact players' motivation to learn and the learning outcomes. In particular, attributes such as fantasy, representation, sensory stimuli, mystery, control, assessment, narrative, realism, adaptivity, interaction, feedback, debriefing, graphics, fun, rules, and goals could support players' learning [42], influence players' cognitive engagement and motivation [20,42,43], and impact players' learning outcomes [20,44].

Despite enjoyment and user experience being essential concepts that are important in fostering learning in serious games, surprisingly few studies [45-50] have measured the dimensions of enjoyment and user experience satisfaction in the serious game context and analyzed how it relates to the learning of players. Giannakos [45], in a study where middle school students played a math game, identified that players' enjoyment had a significant relationship with learning performance. Hamari et al [46], in a study where high school students played an optics game and an engineering dynamics game, identified that engagement had a positive effect on learning, challenge affected learning, and perceived challenge was a strong predictor of learning outcomes. Conversely, they determined that immersion did not have a significant effect on learning. Fokides et al [47], in a study where university students played a history game and a math game, identified that enjoyment, goal clarity, realism, and narration adequacy were the most influential factors shaping user views for serious game learning effectiveness; conversely, they determined that presence, ease of use, and motivation had a marginal effect. Iten and Petko [48], in a study where primary school students (aged 10 to 13 years old) played a game to develop internet skills, identified that enjoyment could encourage players to learn, but it had a small influence on learning gains. In a more recent study with students aged 9 to 12 years old, Iten and Petko [49] identified that self-reported cognitive learning gains were positively correlated with enjoyment. Ebrahimzadeth and Alavi [50], in a study where high school students played a commercial strategy game, identified that enjoyment significantly predicted players' vocabulary learning.

While there is a growing body of literature on the importance of enjoyment and experience satisfaction for players when playing serious video games and their relationship with players learning, currently, there is a dearth of studies that investigated the relationship between these concepts and the learning of young children prior to adolescence (aged between 8 to 10 years) in the context of serious games. Most abovementioned studies [45-50] were evaluated with adolescents (between 10 to 19 years old) [45-49] or young adults [50], and those that included children also included adolescents [48,49], despite the differences in their intellectual and emotional growth [51,52]. The importance of enjoyment and user experience satisfaction and the influence of these concepts on learning could be different for children than for adolescents or young adults [19].

Objective

This study responded to this research gap by examining the relationship between learning, user experience satisfaction, and enjoyment in children aged between 8 to 10 years when playing a serious video game.

Hypotheses

Hypothesis 1: Enjoyment relates positively to children's learning.

Hypothesis 2: Enjoyment influences children's learning.

Hypothesis 3: User experience satisfaction relates positively to children's learning.

Hypothesis 4: User experience satisfaction influences children's learning.

Methods

Overview

This research is part of a comprehensive project to understand the effectiveness of serious video games in children. The video game FoodRateMaster and the research procedure used to obtain the learning data have previously been described in detail [53].

FoodRateMaster

A multidisciplinary team developed FoodRateMaster through an extensive formative research process. FoodRateMaster was designed specifically to teach young children about the suggested ranges for food nutrients to help them determine if they should decrease or maintain their intake of certain foods [53].

FoodRateMaster has 6 levels (Figure 1) that replicate a real food establishment (eg, a food truck, a restaurant, or a grocery store). In each scenario (Figure 2), an evil chef prepares a menu; players, based on the food's nutritional information, determine whether food can be consumed (healthy) or whether its consumption must be reduced (unhealthy). Players use their avatar to take healthy foods (with green labels) and put them into the "keep basket" (green basket) and take unhealthy foods (yellow and red labels) and put them into the "reduce basket" (red basket). Players move the avatar to classify food and avoid obstacles. Maintaining healthy food or reducing unhealthy food allows them to maintain a healthy lifestyle and therefore win points. Additionally, players earn extra points when they correctly replace unhealthy food with healthy food. Otherwise, if players make too many errors, then they do not score enough points to unlock subsequent levels. Finally, a results screen specifies points earned, incorrect food classifications, and position in the global ranking of players (Figure 3).

Several key elements of FoodRateMaster were designed to encourage enjoyment and produce a satisfying user experience [53]: (1) A history of why the players needs to complete the game mission is given using 2 characters: evil chefs and kids who need to be more selective about the food they eat (Figure 4). (2) Narrative, feedback mechanisms, and attractive audio and visual elements were used to encourage the players' immersion, concentration, and enjoyment. (3) A graded task mechanism with a cumulative difficulty curve through different levels was used to adjust the challenge and encourage players' fun through the end of the game (Figure 1). (4) Learning strategies, such as behavioral repetition and substitution (Figure 2) and cognitive restructuring were used to guide knowledge improvement (Figure 5). In addition, in FoodRateMaster, children have to use their creativity and curiosity to figure out how to use the information on nutrition labels and nutrition traffic lights to make better decisions about what foods to eat. An example of these strategies is to use the associations between foods to facilitate and speed up their decision making. (5) Emotional scenes of children were used (eg, when they consume an unhealthy meal, gain weight and can suffer from teasing and bullying). (6) A scoring system and ranking board assessed players' gaming performances.

Figure 1. FoodRateMaster scenario map.

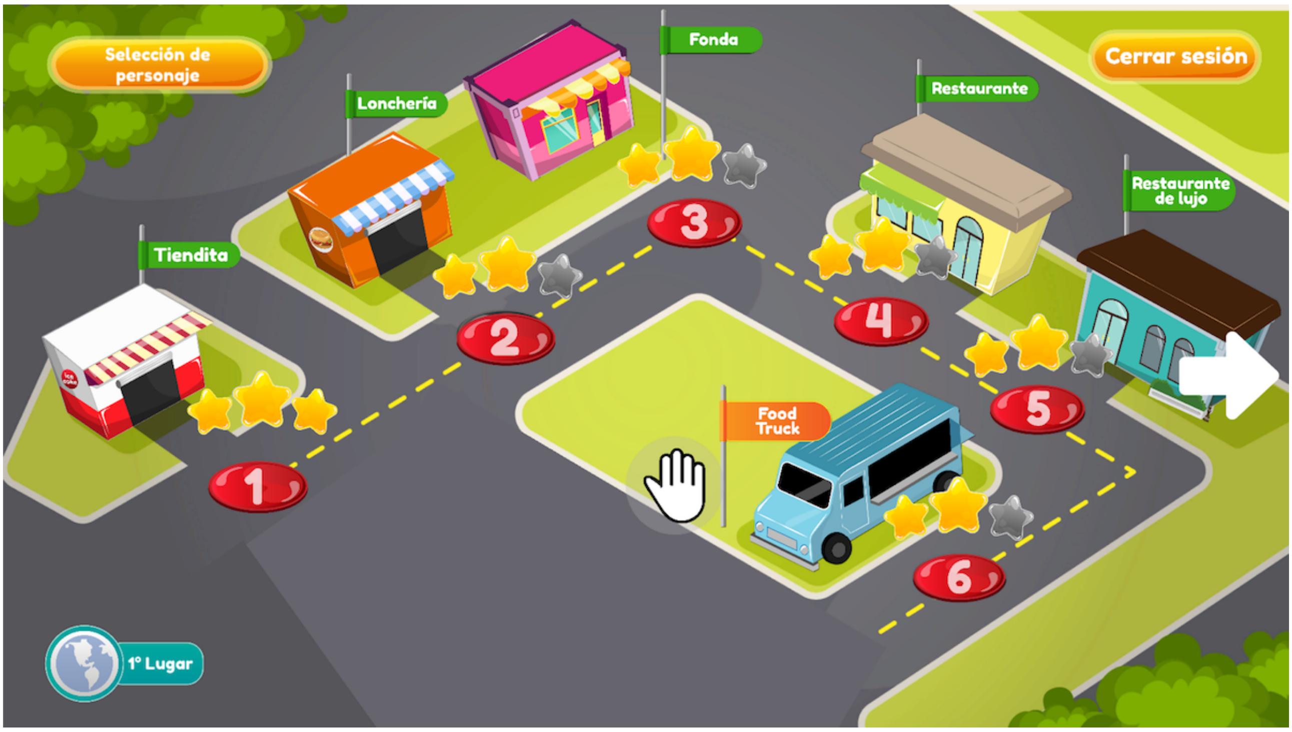


Figure 2. FoodRateMaster game scenario.

Figure 3. FoodRateMaster level results screen.

Figure 4. Story of FoodRateMaster game.

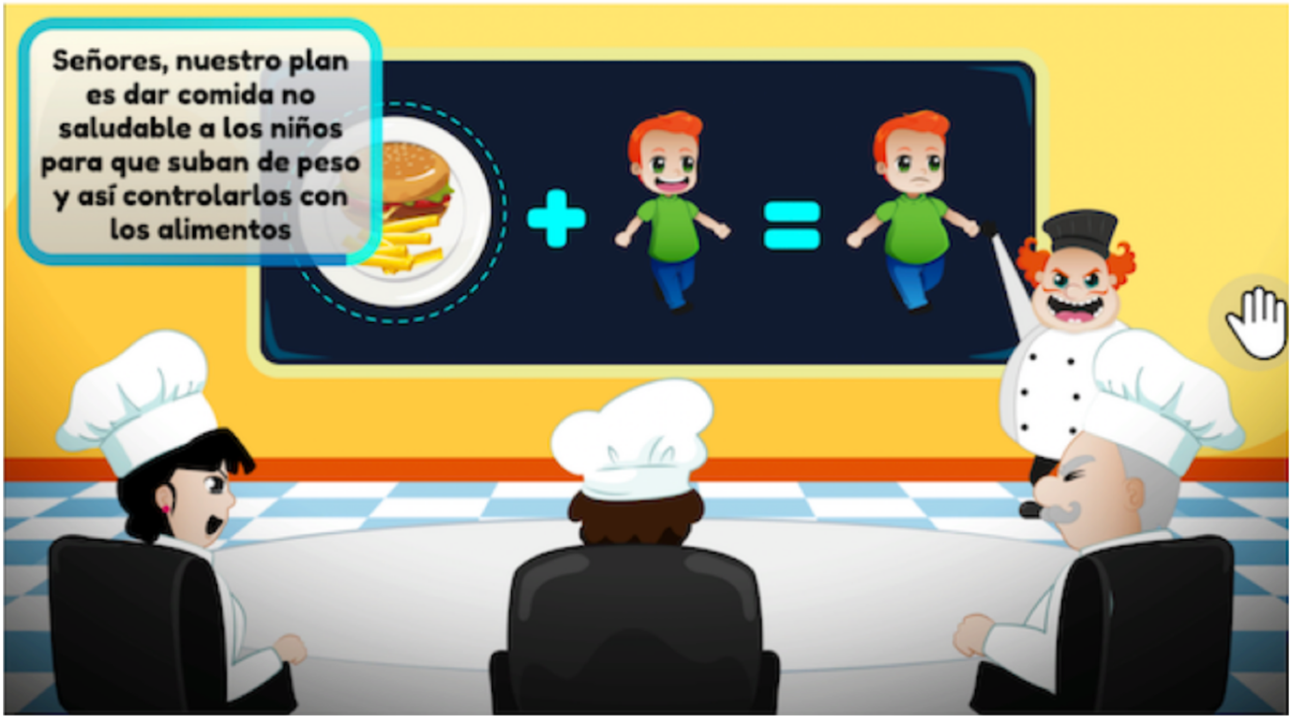


Figure 5. FoodRateMaster explanation of unhealthy food.

Participant Recruitment

Children aged to 10 years (n=62) at a primary school (grades 3, 4, and 5) in Mexico were invited to participate in the pilot study. We conducted a presentation of FoodRateMaster to the children to explain its purpose, characteristics, elements, and how to play it. We obtained written consent from the parents of the children who were interested in participating. Participation was voluntary, and no additional efforts for recruitment were made. The institutional review board of the *Centro de Investigacion Cientifica y de Educacion Superior de Ensenada* (Ensenada Center for Scientific Research and Higher Education) approved this study.

Measures

We used the following 3 questionnaires to collect data: a food knowledge questionnaire [53], the EGameFlow questionnaire [54], and the Game User Experience Satisfaction Scale (GUESS) questionnaire [24].

The food knowledge questionnaire [53] was developed by a multidisciplinary team based on 90 frequently consumed foods in Mexico (49 healthy and 41 unhealthy). A 3-point response format is used to evaluate each food—*healthy* (maintain or increase intake), *unhealthy* (reduce intake), or *I do not know*. The score is the total number of questions answered correctly, with a maximum score of 90.

The EGameFlow questionnaire [54] is a valid and reliable tool for evaluating the level of enjoyment provided by electronic learning games to their user. This scale includes 42 items in 8 subscales [27,54]: *concentration*—games should require concentration, and players should be able to concentrate on the game; *goal clarity*—games should provide players with clear goals at appropriate times; *feedback*—players must receive appropriate feedback about their actions and their progress toward their goals at appropriate times; *challenge*—games should be sufficiently challenging and match players' skill level; *autonomy*—players should feel a sense of control over their actions in the game; *immersion*—players should experience deep but effortless involvement in the game; *social interaction*—games should support and create opportunities for social interaction; *knowledge improvement*—games support and motivate players to acquire, integrate, and apply the knowledge that is taught. The GUESS questionnaire [24] is a valid and reliable tool for evaluating user satisfaction of different video game types by a variety of users with 55 items in 9 subscales: *usability* (or *playability*)—games can be played easily with clear goals or objectives in mind and with minimum interference from the interfaces and controls; *narratives*—the story aspect of the game and its ability to capture players' interest and shape players' emotions; *play engrossment*—games should hold players' attention and interest; *enjoyment*—the pleasure of players that results from playing the game; *creative freedom*—games should foster players' creativity and curiosity; *audio esthetics*—the auditory aspects of the game (eg, sound effects) and how much they enrich the gaming experience; *personal gratification*—players' sense of accomplishment and the desire to succeed and continue playing the game; *social connectivity*—games should facilitate social connections between players; *visual esthetics*—the graphics of the game and how attractive they appear to players. The EGameFlow scale is one of the most widely used questionnaires to evaluate the enjoyment of serious games [55].

We used all subscales for both scales except EGameFlow social interaction and GUESS social connectivity because FoodRateMaster does not include social features. According to Phan et al [24] GUESS subscales that are not applicable (ie, features not included in the video game being evaluated) can be omitted. In both questionnaires, a 5-point Likert-type response format ranging from *strongly disagree* to *strongly* *agree* was used to measure each item. We used only 5 points because having additional points on the response scale provides little additional utility [56] and can influence the item values and the psychological distance [57]. In addition, children can discriminate among 5 response options and do not tend toward the neutral point [58].

Data Collection

In the pretest phase, the participants were assigned unique identifiers and completed the food knowledge questionnaire. During the next 45 days, the children conducted 12 game sessions of at least 15 minutes each with FoodRateMaster (playing time: mean 3.5 hours, SD 0.8). In the posttest phase (the day after the last game session), children completed the food knowledge questionnaire (again), the EGameFlow questionnaire [54], and the GUESS questionnaire [24]. The children took approximately 15 minutes to complete each questionnaire. Initially, we conducted a pilot test of these questionnaires and identified that children did not have any problem when answering them.

Data Analysis

Data analysis was carried out using the SPSS statistical software (version 26; IBM Corp). In all analyzes, the statistical significance was 5% ($P<.05$). A paired two-tailed t test was conducted to examine the difference between pretest and posttest knowledge scores. Pearson correlation was conducted to examine the relationships of enjoyment and user experience satisfaction, including their subscales, with posttest knowledge scores. We also performed standard multiple regression on EGameFlow and the GUESS scores to identify if their subscales could predict the results of the posttest children's knowledge and to examine their correlations. Separate multiple linear regressions were calculated to predict the posttest children's knowledge based on (1) EGameFlow and (2) GUESS subscales. Preliminary analyses were performed to ensure that the assumptions of linear regression analysis were not violated (the criteria are included in parentheses) [59,60]: linearity, normality, nonzero variance (variance >0), no data outliners (standardized residual values for each participant from −3.29 to 3.29), collinearity (for all variables, variance inflation factor is <10 or tolerance >0.1), and independent errors (Durbin-Watson values from 1.5 to 2.5).

Results

Data from 2 children were excluded from the analysis (ie, these 2 students did not complete all the questionnaires); thus, the study completion rate was 97% (60/62): 29 girls and 31 boys (age: mean 9, SD 0.8 years).

The mean pretest and posttest knowledge scores were 56.95 (SD 10.71, range 30-75) and 67.88 (SD 10.71, range 30-85), respectively. In addition, the mean enjoyment and user experience satisfaction scores were 4.16 (SD 0.55, range 2.28-5.00) and 4.29 (SD 0.54, range 2.79-5.00), respectively. Knowledge scores were significantly lower in the pretest than in the posttest (mean difference −11.12, SD 8.27; $t_{59}=-10.41$, $P<.001$). Thus, children increased their nutritional knowledge after game play.

We found associations (Table 1) between posttest knowledge scores and enjoyment (EGameFlow scale: $r_{58}=0.36$, $P=.005$), as well as between posttest knowledge scores and user experience satisfaction (GUESS scale: $r_{58}=0.34$, $P=.008$). Additionally, we found associations between posttest knowledge and the following EGameFlow subscales: challenge ($r_{58}=0.38$, $P=.003$), knowledge improvement ($r_{58}=0.38$, $P=.003$), and goal clarity ($r_{58}=0.29$, $P=.02$). Also, associations were found between posttest knowledge and the following GUESS subscales: narrative ($r_{58}=0.35$, $P=.006$), creative freedom ($r_{58}=0.26$, $P=.04$), and visual esthetics ($r_{58}=0.32$, $P=.01$).

Both models fulfilled all the assumptions of linear regression analysis. The regression results indicated that the model based on the EGameFlow scale explained 27% of the variance (large effect [61]) and that the model was a significant predictor of children's learning ($F_{7,52}=2.74$, $P=.02$). Table 2 describes the contribution of each subscale to the total explained variance; only 2 subscales had significant contributions, namely, challenge ($\beta=0.40$, $t_{52}=2.17$, $P=.04$) and knowledge improvement ($\beta=.29$, $t_{52}=2.06$, $P=.04$). Meanwhile, the regression results indicated that the model based on the GUESS scale explained 26% of the variance (large effect) and that the model was a significant predictor of children's learning ($F_{8,51}=2.20$, $P=.04$). However, as shown in Table 2, none of the subscales had a significant contribution.

Table 1. Bilateral correlations of children's posttest knowledge with EGameFlow and Guess scales and subscales.

EGameFlow and GUESS subscales	Posttest knowledge score	
	r	*P* value (bilateral)
EGameFlow (Enjoyment[a])	**0.36**	**.005** [b]
Concentration	0.09	.49
Goal clarity	0.29	.02[b]
Feedback	0.17	.19
Challenge	0.38	.003[b]
Autonomy	0.01	.93
Immersion	0.06	.65
Knowledge improvement	0.38	.003[b]
GUESS (User experience satisfaction[c])	**0.27**	**.04** [b]
Usability/playability	0.21	.11
Narratives	0.35	.006[b]
Play engrossment	−0.15	.25
Enjoyment	0.09	.47
Creative freedom	0.26	.04[b]
Audio esthetics	0.09	.45
Personal gratification	0.15	.24
Visual esthetics	0.32	.01[b]

[a]Enjoyment as a whole (mean of subscale scores).

[b]*P*<.05.

[c]User experience satisfaction as a whole (mean of subscale scores).

Table 2. Model coefficients for EGameFlow and GUESS subscale ratings with posttest knowledge as the dependent variable.

EGameFlow and GUESS subscales	β	*t* value[a]	*P* value
EGameFlow scale—Enjoyment			
Concentration	0.03	0.24	.81
Goal clarity	0.14	0.89	.38
Feedback	0.13	−0.72	.48
Challenge	0.40	2.17	.04[b]
Autonomy	0.08	−0.58	.56
Immersion	0.19	−1.30	.20
Knowledge improvement	0.29	2.06	.04[b]
GUESS scale—User experience satisfaction			
Usability/playability	0.30	1.41	.16
Narratives	0.26	1.69	.10
Play engrossment	−0.23	−1.79	.08
Enjoyment	0.02	0.12	.91
Creative Freedom	0.22	1.42	.16
Audio esthetics	−0.34	−1.93	.06
Personal gratification	−0.21	−0.99	.32
Visual esthetics	0.22	1.19	.24

[a]Model *df*=52 for EGameFlow and *df*=51 for GUESS.

[b]*P*<.05.

Discussion

Principal Results

This study investigated the relationships of enjoyment and user experience satisfaction with children's learning while playing a serious video game. As far as we know, this is the first paper analyzing these relationships in children aged 8 to 10 years. Overall, the study's findings suggest that serious games can provide enjoyment and satisfactory user experiences to children, as demonstrated by the levels of enjoyment and user experience satisfaction. The results specifically showed significant associations for children's enjoyment and user experience satisfaction with their learning when playing serious video games. In addition, the ratings given by children on subscales of the questionnaires used to assess enjoyment and user experience satisfaction were significant predictors of their learning. Challenge and knowledge improvement (EGameFlow) were correlated with children's learning and contributed to predicting it. Meanwhile, narrative, creative freedom, and visual esthetics (GUESS) were correlated with children's learning; however, did not contribute toward predicting learning.

Comparison With Prior Work

Associating the findings with the results of previous research is a difficult task, given that our study examined the influence of enjoyment and user experience on the learning of young children (aged between 8 to 10 years old), in contrast to the existing studies that involved adolescents or young adults [45-47]. Even so, the findings that enjoyment and user experience were associated with learning and can help to predict player learning provide further support to findings in previous studies [45-47,49,50]. In addition, this is the first study (as far as we know) that analyzed the relationship of these factors in the context of a serious game for health.

Enjoyment and Game-based Learning

Hypotheses 1 and 2, which emerged from the expectation that enjoyment positively correlates to children's learning and can predict it, were confirmed (hypothesis 1: r_{58}=0.36, *P*=.005; hypothesis 2: $F_{7,52}$=2.74, *P*=.02). These results are in line with previous findings that enjoyment had a significant relationship with learning performance [45-47,49] and that enjoyment can predict learning [50], and in addition, that this factor significantly contributed to the effectiveness of serious games [4,17].

Using the EGameFlow scale, challenge and knowledge improvement were found to correlate with and contribute to predicting children's learning. Regarding challenge, this finding was anticipated since the relationship between challenge and skill determines players' engagement level during gameplay [62]. Thus, the optimal amount of difficulty should match user's abilities to the skills required to accomplish each goal [27]. This result supports the findings of several studies that have indicated that challenge is a critical factor in the success of serious games [20,29,34,43] and related to players' learning [46,50]. Regarding knowledge improvement, serious games should incorporate strategies to provide knowledge that players need to learn and integrate to fulfill the goals of the game [54]. Our results supports those of Ebrahimzadeth and Alavi [50], who identified that knowledge improvement is positively correlated with players' learning. In addition, goal clarity correlated with children's learning. This results add support to the findings of studies [43,47] that stated that goal clarity enhances the engagement levels of players and influences learning effectiveness. However, this result contrasts with the findings of other studies [46,47,50] that identified that goal clarity does not influence the players' learning.

Concentration (P=.49), feedback (P=.19), autonomy (P=.93), and immersion (P=.65) were not correlated with children's learning, nor did they predicting it. These results may be related to how children perceive the feedback of serious games, their need for autonomy and cognition, and their sensation seeking [43]. This result gives further support to those of previous studies [46,47,50], noting that concentration, feedback and immersion do not influence players' learning. However, this result contrasts with findings in other studies stating that concentration is essential for learning in serious games [18,27], that feedback influences the learning of players [18,20,36,43,44], and that autonomy is an important factor in the effectiveness of serious games [20,29,43] and is correlated with players' learning [50]. Finally, the actual impact of immersion on learning outcomes seems more complicated in the use of video games for educational purposes [63]. While some studies [31,50,64] have noted that immersion is an important factor in the effectiveness of serious games [31] and is related to players' learning [50,64], other studies [63,65] have identified that immersion can limit players' learning because players can ignore the educational targets of video games due to the increased cognitive load.

User Experience Satisfaction and Game-based Learning

Hypotheses 3 and 4, emerging from the expectation that user experience satisfaction positively correlates to and can predict children's learning were confirmed by the data analysis (hypothesis 3: r_{58}=0.27, P=.04; hypothesis 4: $F_{8,51}$=2.20, P=.04). Comparing these findings with those of previous research is difficult due to most existing studies focusing on using the GUESS scale to evaluate serious games (eg, [66,67]) and (as far we know) no previous studies have addressed the relationship between the GUESS scale and subscales with players' learning.

Using GUESS, narrative, creative freedom, and visual esthetics correlated with children's learning, but none contributed toward predicting children's learning. Narration facilitates situated cognition by immersing players in a setting that frames knowledge [68]. Our results support those of studies that noted the impact of a narrative on learning [17,21,29,43,44,69] and that identified that a narrative is positively correlated with players' learning [47,50]. Regarding creative freedom, playing is a creative process by itself that can foster new and novel associations between ideas, objects, and behaviors [70]. In this way, serious games provide a safe environment for active, critical, and creative learning, allowing users to explore skills, methods, and concepts [71]. Previous studies [72,73] have reported the usefulness of serious games for ideation; however, no previous studies have analyzed the correlation between creative freedom and players' learning, as far as we know. In addition, our results give further support to those of previous studies noting the influence of visual esthetics on the acceptability of serious games and the learning of players [17,20,29,33,42,44] and reporting a statistical association of the esthetic presentation of the game upon the learning outcome [74].

In this study, usability (or playability) (P=.11), play engrossment (P=.25), enjoyment (P=.47), audio esthetics (P=.45), and personal gratification (P=.24) did not reach a statically significant correlation with children's learning, nor did they contribute to predicting it. Our results support the results of Fokides et al [47], who found that there was no significant relationship between usability (or playability) and the learning of players and that it only influenced players' enjoyment. However, our results are in contrast to the findings of Tolentino et al [35], who stated that serious video games should have a high level of usability. Player engrossment is similar to the immersion subscale of the EGameFlow scale, which was not associated with learning. The absence of correlation between the enjoyment subscale of the GUESS scale and children's learning is surprising because there was a correlation between the score of the EGameFlow scale and children's learning. We consider that this result is due to the enjoyment subscale of the GUESS scale and does not include all the factors that influence enjoyment considered in the EGameFlow scale. Finally, although previous research also noted the importance of audio esthetics [20,29,75] and personal gratification [29] to generate motivation and immersion in serious games, no previous studies have analyzed the correlation between these subscales and the learning of players, as far as we know.

Limitations and Future Work

This study faced some limitations. First, the self-report measure of enjoyment and user experience satisfaction can be highly sensitive to the respondent's comprehension and willingness to provide honest answers. Second, the results should be interpreted with caution since a small sample of children participated in the study. The sample size could have been larger, allowing for more confidence in the robustness of the statistical analysis and results. However, given that our main objective was to explore the relationship of children's learning with enjoyment and user experience satisfaction, we believe that our results provide a good starting point for researchers to explore the design of serious video games for children. Third, due to space and time constraints, children who were not from the same class attended at the same time to play on the play stations. This situation might have affected the participant behaviors (ie, higher or lower performance in the game). Fourth, some students were more familiar with video games, which might have affected their performance. Fifth, the frequency and duration of the sessions were predefined. In real environments, it is common for game sessions to be longer and more frequent, which could accelerate the rate at which players become bored with the video game and can even generate video gaming burnout [76], consequently affecting learning, enjoyment, and user' satisfaction. Sixth, the students also had a limited age range, limiting the generalizability of the findings. Finally, the children played only one serious game. To overcome these issues, we suggest similar studies be carried out with students of different age groups and levels of experience with video games, using several different serious video games, and where players can carry out their sessions at their own pace.

For future work, we plan to carry out a comprehensive study with users in a broad age range to provide generalizable results. In addition, based on the insight that the influence of enjoyment and user experience satisfaction could depend on the age of players, in a future study, we also plan to perform age-based analysis to assess the importance of age and determine its association with the relationship of enjoyment and user experience satisfaction with learning, as well as whether there is variability between different subscales among users of different ages.

Conclusions

In summary, the reported findings from this study can be useful to better understand the correlation and influence of enjoyment and user experience on children's learning when playing serious games. These findings should be considered in the design stages during the development of serious video games for children and suggest that, because children's enjoyment and user experience satisfaction are significantly associated with their learning, and enjoyment and user experience satisfaction are significant predictors of children's learning while playing a serious video game, when serious video games are designed for children, it is essential to implement elements that foster enjoyment and user experience satisfaction. In particular, video games should be sufficiently challenging, match children's skills, and include elements supporting and motivating players to acquire and apply the learned knowledge. Clear goals, narrative, creative freedom, and visual esthetics are positively correlated with children's learning; therefore, designers of serious video games for children should dedicate efforts to providing clear goals and producing elements that capture players' interest, shape players' emotions, and foster their creativity and curiosity, as well as include attractive graphical elements in the game.

Acknowledgments

We thank the *Consejo Nacional de Ciencia y Tecnología* (CONACYT, Mexico) for financial support (grant number PDCPN-2015-824). We thank the *Colegio Inglés de Occidente* for facilitating the conduct of this study and all of the teachers, students, and parents who participated.

Authors' Contributions

IEE-C designed the study, conducted data collection and processing, wrote the first draft, incorporated changes from all coauthors, and completed the data analyses. EEP-B developed FoodRateMaster and contributed to data collection. JM-M and HP-E assisted with the analysis and interpretation of the data and contributed to editing the manuscript.

Conflicts of Interest

None declared.

References

1. Girard C, Ecalle J, Magnan A. Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. Journal of Computer Assisted Learning 2013 Jun 13;29(3):207-219. [doi: 10.1111/j.1365-2729.2012.00489.x]
2. Thompson D, Baranowski T, Buday R, Baranowski J, Thompson V, Jago R, et al. Serious Video Games for Health How Behavioral Science Guided the Development of a Serious Video Game. Simul Gaming 2010 Aug 01;41(4):587-606. [doi: 10.1177/1046878108328087] [Medline: 20711522]
3. Laamarti F, Eid M, El Saddik A. An Overview of Serious Games. International Journal of Computer Games Technology 2014;2014:1-15. [doi: 10.1155/2014/358152]
4. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
5. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 2013;105(2):249-265. [doi: 10.1037/a0031311]
6. Boyle EA, Hainey T, Connolly TM, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education 2016 Mar;94:178-192. [doi: 10.1016/j.compedu.2015.11.003]
7. Zhonggen Y. A Meta-Analysis of Use of Serious Games in Education over a Decade. International Journal of Computer Games Technology 2019 Feb 03;2019:1-8. [doi: 10.1155/2019/4797032]
8. Yanti PP, Rosmansyah Y, Dabarsyah B. Serious Games for Children: A Systematic Literature Review. In: International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS).: IEEE; 2019 Presented at: 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS); 24-25 Oct. 2019; Jakarta, Indonesia, Indonesia p. 79-84. [doi: 10.1109/ICIMCIS48181.2019.8985230]

9. Parisod H, Pakarinen A, Kauhanen L, Aromaa M, Leppänen V, Liukkonen TN, et al. Promoting Children's Health with Digital Games: A Review of Reviews. Games Health J 2014 Jun;3(3):145-156. [doi: 10.1089/g4h.2013.0086] [Medline: 26196173]
10. Holtz BE, Murray K, Park T. Serious Games for Children with Chronic Diseases: A Systematic Review. Games Health J 2018 Oct;7(5):291-301. [doi: 10.1089/g4h.2018.0024] [Medline: 29957082]
11. Hainey T, Connolly TM, Boyle EA, Wilson A, Razak A. A systematic literature review of games-based learning empirical evidence in primary education. Computers & Education 2016 Nov;102:202-223. [doi: 10.1016/j.compedu.2016.09.001]
12. Chow CY, Riantiningtyas RR, Kanstrup MB, Papavasileiou M, Liem GD, Olsen A. Can games change children's eating behaviour? A review of gamification and serious games. Food Quality and Preference 2020 Mar;80:103823. [doi: 10.1016/j.foodqual.2019.103823]
13. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for Health for Children-Current Status and Needed Research. Games Health J 2016 Feb;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
14. Meluso A, Zheng M, Spires HA, Lester J. Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. Computers & Education 2012 Sep;59(2):497-504. [doi: 10.1016/j.compedu.2011.12.019]
15. Lu AS, Kharrazi H, Gharghabi F, Thompson D. A Systematic Review of Health Videogames on Childhood Obesity Prevention and Intervention. Games Health J 2013 Jun;2(3):131-141 [FREE Full text] [doi: 10.1089/g4h.2013.0025] [Medline: 24353906]
16. Lu AS, Kharrazi H. A State-of-the-Art Systematic Content Analysis of Games for Health. Games Health J 2018 Feb;7(1):1-15. [doi: 10.1089/g4h.2017.0095] [Medline: 29293368]
17. Shen C, Wang H, Ritterfeld U. Chapter 4. Serious Games and Seriously Fun Games: Can They Be One and the Same? In: Ritterfeld U, Cody M, Vorderer P, editors. Serious Games Mechanisms and Effects. New York, NY: Taylor & Francis; 2009:1-15.
18. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul (Lond) 2017 Feb 04;2:3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
19. M.Nazry N, Romano DM. Mood and learning in navigation-based serious games. Computers in Human Behavior 2017 Aug;73:596-604. [doi: 10.1016/j.chb.2017.03.040]
20. Wilson KA, Bedwell WL, Lazzara EH, Salas E, Burke CS, Estock JL, et al. Relationships Between Game Attributes and Learning Outcomes: Review and Research Proposals. Simulation & Gaming 2008 May 07;40(2):217-266. [doi: 10.1177/1046878108321866]
21. Crutzen R, van 't Riet J, Short CE. Enjoyment: A Conceptual Exploration and Overview of Experimental Evidence in the Context of Games for Health. Games Health J 2016 Feb;5(1):15-20. [doi: 10.1089/g4h.2015.0059] [Medline: 26699455]
22. Schaffer O, Fang X. Digital Game Enjoyment: A Literature Review. In: HCI in Games.: Springer International Publishing; 2019 Presented at: First International Conference, HCI-Games 2019; July 26-31, 2019; Orlando, FL, USA p. 191-214. [doi: 10.1007/978-3-030-22602-2_16]
23. González Sánchez JL, Gutiérrez Vela FL, Montero Simarro F, Padilla-Zea N. Playability: analysing user experience in video games. Behaviour & Information Technology 2012 Oct;31(10):1033-1054. [doi: 10.1080/0144929X.2012.710648]
24. Phan MH, Keebler JR, Chaparro BS. The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). Hum Factors 2016 Dec;58(8):1217-1247. [doi: 10.1177/0018720816669646] [Medline: 27647156]
25. Wiemeyer J, Nacke L, Moser C, Mueller F. Player Experience. In: Dörner R, Göbel S, Effelsberg W, Wiemeyer J, editors. Serious Games: Foundations, Concepts and Practice. Switzerland: Springer; Aug 2016:243-271.
26. Moizer J, Lean J, Dell'Aquila E, Walsh P, Keary A, O'Byrne D, et al. An approach to evaluating the user experience of serious games. Computers & Education 2019 Jul;136:141-151. [doi: 10.1016/j.compedu.2019.04.006]
27. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Computers in Entertainment 2005 Jul 01;3(3):1-24. [doi: 10.1145/1077246.1077253]
28. Wang Y, Rajan P, Sankar CS, Raju PK. Let Them Play: The Impact of Mechanics and Dynamics of a Serious Game on Student Perceptions of Learning Engagement. IEEE Transactions on Learning Technologies 2017 Oct 1;10(4):514-525. [doi: 10.1109/tlt.2016.2639019]
29. Yen-Ru S, Ju-Ling S. Game Factors and Game-Based Learning Design Model. International Journal of Computer Games Technology 2015;2015:1-11. [doi: 10.1155/2015/549684]
30. Baranowski T, Buday R, Thompson D, Lyons EJ, Lu AS, Baranowski J. Developing Games for Health Behavior Change: Getting Started. Games Health J 2013 Aug;2(4):183-190 [FREE Full text] [doi: 10.1089/g4h.2013.0048] [Medline: 24443708]
31. Annetta LA. The "I's" have it: A framework for serious educational game design. Review of General Psychology 2010;14(2):105-113. [doi: 10.1037/a0018985]
32. Wrzesien M, Alcañiz Raya M. Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. Computers & Education 2010 Aug;55(1):178-187. [doi: 10.1016/j.compedu.2010.01.003]
33. Wilson AS, Broadbent C, McGrath B, Prescott J. Chapter 23. Factors Associated with Player SatisfactionEducational Value of Serious Games. In: Ma M, Oikonomou A, editors. Serious Games and Edutainment Applications Volume II. Switzerland: Springer International Publishing; 2017:513-535.

34. Nagle A, Novak D, Wolf P, Riener R. The effect of different difficulty adaptation strategies on enjoyment and performance in a serious game for memory training. In: IEEE International Conference on Serious Games and Applications for Health (SeGAH). 2014 Presented at: 2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH); 2014; Rio de Janeiro, Brazil p. 1-8. [doi: 10.1109/segah.2014.7067088]
35. Pino Tolentino G, Battaglini C, Ventura Pereira AC, Jacó de Oliveria R, Mendes de Paula MG. Usability of Serious Games for Health. In: Conference on Games and Virtual Worlds for Serious Applications, VS-GAMES.: IEEE; 2011 Presented at: 2011 Third International Conference on Games and Virtual Worlds for Serious Applications; 4-6 May 2011; Athens, Greece p. 172-175. [doi: 10.1109/vs-games.2011.33]
36. Ifenthaler D, Eseryel D, Ge X. Assessment for Game-Based Learning. In: Ifenthaler D, Eseryel D, Ge X, editors. Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives. New York, NY: Springer; 2012:1-8.
37. Schneider KL, Ferrara J, Lance B, Karetas A, Druker S, Panza E, et al. Acceptability of an Online Health Videogame to Improve Diet and Physical Activity in Elementary School Students:. Games Health J 2012 Aug;1(4):262-268 [FREE Full text] [doi: 10.1089/g4h.2012.0009] [Medline: 24761317]
38. Eckardt L, Robra-Bissantz S. EGameFlow in a Serious Game: Gaming Experience with the Same Game Design but Different Learning Content. In: Gaming Experience in a Serious Game with Different Learning Content. EGameFlow in a Serious Game: Gaming Experience with the Same Game Design but Different Learning Content. Twenty-Third Pac Asia Conf Inf Syst Xi'an, China; 2019 Presented at: Pacific Asia Conference on Information Systems (PACIS 2019); 8-12 July, 2019; Xi'an, China p. 1-8.
39. Buday R, Baranowski T, Thompson D. Fun and Games and Boredom. Games Health J 2012 Aug;1(4):257-261 [FREE Full text] [doi: 10.1089/g4h.2012.0026] [Medline: 24761316]
40. Kelley C, Wilcox L, Ng W, Schiffer J, Hammer J. Design Features in Games for Health: Disciplinary and Interdisciplinary Expert Perspectives. DIS (Des Interact Syst Conf) 2017 Jun;2017:69-81 [FREE Full text] [doi: 10.1145/3064663.3064721] [Medline: 28868523]
41. Rieber LP, Noah D. Games, simulations, and visual metaphors in education: antagonism between enjoyment and learning. Educational Media International 2008 Jun;45(2):77-92. [doi: 10.1080/09523980802107096]
42. Huang WD, Johnson TE, Han SC. Impact of online instructional game features on college students' perceived motivational support and cognitive investment: A structural equation modeling study. The Internet and Higher Education 2013 Apr;17:58-68. [doi: 10.1016/j.iheduc.2012.11.004]
43. Alexiou A, Schippers MC. Digital game elements, user experience and learning: A conceptual framework. Education and Information Technologies 2018 May 17;23:2545-2567. [doi: 10.1007/s10639-018-9730-6]
44. Ravyse WS, Seugnet Blignaut A, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virtual Reality 2016 Sep 20;21:31-58. [doi: 10.1007/s10055-016-0298-4]
45. Giannakos MN. Enjoy and learn with educational games: Examining factors affecting learning performance. Computers & Education 2013 Oct;68:429-439. [doi: 10.1016/j.compedu.2013.06.005]
46. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. Computers in Human Behavior 2016 Jan;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
47. Fokides E, Atsikpasi P, Kaimara P, Deliyannis I. Factors Influencing the Subjective Learning Effectiveness of Serious Games. Journal of Information Technology Education: Research 2019;18:437-466. [doi: 10.28945/4441]
48. Iten N, Petko D. Learning with serious games: Is fun playing the game a predictor of learning success? British Journal of Educational Technology 2016 Jan;47(1):151-163. [doi: 10.1111/bjet.12226]
49. Imlig-Iten N, Petko D. Comparing Serious Games and Educational Simulations: Effects on Enjoyment, Deep Thinking, Interest and Cognitive Learning Gains. Simulation & Gaming 2018 Aug 01;49(4):401-422. [doi: 10.1177/1046878118779088]
50. Ebrahimzadeh M, Alavi S. Motivating EFL students: E-learning enjoyment as a predictor of vocabulary learning through digital video games. Cogent Education 2016 Nov 14;3(1255400):1-14. [doi: 10.1080/2331186x.2016.1255400]
51. Gilmore K, Meersand P. Chapter 6.The Latency Phase. In: Normal Child And Adolescent Development: A Psychodynamic Primer. Washington, DC: American Psychiatric Publishing; 2014:364.
52. Colarusso CA. Child and Adult Development: A Psychoanalytic Introduction for Clinicians. New York: Springer Science & Business Media; 1992:220.
53. Espinosa-Curiel IE, Pozas-Bogarin EE, Lozano-Salas JL, Martínez-Miranda J, Delgado-Pérez EE, Estrada-Zamarron LS. Nutritional Education and Promotion of Healthy Eating Behaviors Among Mexican Children Through Video Games: Design and Pilot Test of FoodRateMaster. JMIR Serious Games 2020 Apr 13;8(2):e16431 [FREE Full text] [doi: 10.2196/16431] [Medline: 32281539]
54. Fong-Ling F, Rong-Chang S, Sheng-Chin Y. EGameFlow: A scale to measure learners' enjoyment of e-learning games. Computers & Education 2009 Jan;52(1):101-112. [doi: 10.1016/j.compedu.2008.07.004]
55. Petri G, Wangenheim CGV. How to Evaluate Educational Games: a Systematic Literature Review. Journal of Universal Computer Science 2016;22(7):992-1021. [doi: 10.3217/jucs-022-07-0992]
56. Lissitz RW, Green SB. Effect of the number of scale points on reliability: A Monte Carlo approach. Journal of Applied Psychology 1975 Feb;60(1):10-13. [doi: 10.1037/h0076268]

57. Wakita T, Ueshima N, Noguchi H. Psychological Distance Between Categories in the Likert Scale: Comparing Different Numbers of Options. Educational and Psychological Measurement 2012 Jan 12;72(4):533-546. [doi: 10.1177/0013164411431162]
58. Adelson JL, McCoach DB. Measuring the Mathematical Attitudes of Elementary Students: The Effects of a 4-Point or 5-Point Likert-Type Scale. Educational and Psychological Measurement 2010 Apr 13;70(5):796-807. [doi: 10.1177/0013164410366694]
59. Verma J, Abdel-Salam G. Testing Statistical Assumptions in Research. Hoboken, NJ: John Wiley & Sons; Mar 2019:978.
60. Weinberg SL, Abramowitz SK. Statistics using SPSS: An integrative approach, third edition. New York, NY: Cambridge University Press; 2015:595.
61. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155]
62. Csikszentmihalyi M. Beyond Boredom and Anxiety: Experiencing Flow in Work and Play. San Francisco, California: Jossey-Bass; 2000:272.
63. Cheng M, She HC, Annetta L. Game immersion experience: its hierarchical structure and impact on game-based science learning. Journal of Computer Assisted Learning 2015 Jun;31(3):232-253. [doi: 10.1111/jcal.12066]
64. Georgiou Y, Kyza EA. Relations between student motivation, immersion and learning outcomes in location-based augmented reality settings. Computers in Human Behavior 2018 Dec;89:173-181. [doi: 10.1016/j.chb.2018.08.011]
65. Cheng MT, Su T, Huang WY, Chen JH. An educational game for learning human immunology: What do students learn and how do they perceive? British Journal of Educational Technology 2014 Sep;45(5):820-833. [doi: 10.1111/bjet.12098]
66. Smith PA, Dombrowski M, Buyssens R, Barclay P. Usability testing games for prosthetic training. In: IEEE International Conference on Serious Games and Applications for Health (SeGAH).: IEEE; 2018 Presented at: 018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH); 16-18 May 2018; Vienna, Austria p. 1-7. [doi: 10.1109/SeGAH.2018.8401376]
67. Espinosa-Curiel IE, Martínez-Rosas MJ, del Hoyo-Ceja JM, Delgado-Pérez EE, Pozas-Bogarin EE. HealthyLunch: A Serious Game for Educating and Promoting the Intake of the Recommended Number of Daily Servings Among Children. In: Games and Learning Alliance. Cham: Springer; 2019 Presented at: International Conference on Games and Learning Alliance 2019 (GALA 2019); November 27–29; Athens, Greece p. A-361. [doi: 10.1007/978-3-030-34350-7_34]
68. Van Eck R. Digital Game-Based Learning: It's Not Just the Digital Natives Who Are Restless. Educause Review 2006;41(2):16-30.
69. Calderón A, Ruiz M. A systematic literature review on serious games evaluation: An application to software project management. Computers & Education 2015 Sep;87:396-422. [doi: 10.1016/j.compedu.2015.07.011]
70. Bowman N, Kowert R, Ferguson C. Chapter 2 - The Impact of Video Game Play on Human (and Orc) Creativity. In: Green GP, Kaufman JC, editors. Video Games and Creativity. London, UK: Academic Press; 2015:39-60.
71. Protopsaltis A, Pannese L, Hetzner S, Pappa D, De Freitas S. Creative Learning with Serious Games. International Journal of Emerging Technologies in Learning 2010 Dec;5:4-6. [doi: 10.3991/ijet.v5s3.1495]
72. Agogué M, Levillain K, Hooge S. Gamification of Creativity: Exploring the Usefulness of Serious Games for Ideation. Creativity and Innovation Management 2015 Sep;24(3):415-429. [doi: 10.1111/caim.12138]
73. Hauge JB, Duin H, Thoben KD. The Evaluation of Serious Games Supporting Creativity through Student Labs. In: Serious Games Development and Applications. Berlin, Heidelberg: Springer; 2013 Presented at: 4th International Conference, SGDA 2013; September 25-27, 2013; Trondheim, Norway p. 188-199. [doi: 10.1007/978-3-642-40790-1_18]
74. Martin MW, Shen Y. The Effects of Game Design on Learning Outcomes. Computers in the Schools 2014 Apr 18;31(1-2):23-42. [doi: 10.1080/07380569.2014.879684]
75. Jabbar AIA, Felicia P. Gameplay Engagement and Learning in Game-Based Learning. Review of Educational Research 2015 Dec 1;85(4):740-779. [doi: 10.3102/0034654315577210]
76. Badrinarayanan VA, Sierra JJ, Taute HA. Determinants and Outcomes of Online Brand Tribalism: Exploring Communities of Massively Multiplayer Online Role Playing Games (MMORPGs). Psychology and Marketing 2014 Oct;31(10):853-870. [doi: 10.1002/mar.20739]

Abbreviations

GUESS: Game User Experience Satisfaction Scale
STEM: science, technology, engineering, math

Edited by G Eysenbach; submitted 25.06.20; peer-reviewed by S Dick, K Murray, D Drummond; comments to author 11.07.20; revised version received 21.08.20; accepted 01.09.20; published 17.09.20.

Please cite as:
Espinosa-Curiel IE, Pozas-Bogarin EE, Martínez-Miranda J, Pérez-Espinosa H
Relationship Between Children's Enjoyment, User Experience Satisfaction, and Learning in a Serious Video Game for Nutrition Education: Empirical Pilot Study
JMIR Serious Games 2020;8(3):e21813
URL: http://games.jmir.org/2020/3/e21813/
doi:10.2196/21813
PMID:32940609

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org